The Dissertation Committee for Xiaoxia Cui Certifies that this is the approved version of the following dissertation:

RNA/protein interactions during group II intron splicing and toward group II intron targeting in mammalian cells

Committee:

Alan M. Lambowitz, Supervisor

Karen Browning

Arlen Johnson

Robert Krug

Paul Macdonald

RNA/protein interactions during group II intron splicing and toward group II intron targeting in mammalian cells

by

Xiaoxia Cui, B.S., M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2005

Dedication

To Grandma, Dad, Mom, and Fei

Acknowledgements

First I thank my supervisor, Dr. Alan M. Lambowitz, for his guidance, inspiration, and patience throughout the years, and his dedication to science will influence me for a lifetime. I would also like to thank members of my dissertation committee, Drs. Karen Browning, Arlen Johnson, Robert Krug, and Paul Macdonald, for their time, advice, and support. I am grateful to Dr. Vytas A. Bankaitis, my master's thesis advisor, whose encouragement and belief in me helped me through difficult times.

I thank all members of the Lambowitz Lab, past and present, for their help and friendship, especially Drs. Georg and Sabine Mohr, Roland Saldanha, Manabu Matsuura, Huatao Guo, Karolina Briegel, Kazuo Watanabe, Forrest Blocker, Jiri Perutka, Ms. Qin Wang, Ms. Junhua Zhao, Mr. Mike Karberg, and Mr. Lawrence Manzano.

Finally, I thank my family for their ultimate support, especially, Dad, who led me into life sciences, and whose expectations have been my driving forces, Fei, my husband, without whose love and encouragement, I would not have made it this far, and my brother Feng, who even did some minipreps for me.

RNA/protein interactions during group II intron splicing and toward group II intron targeting in mammalian cells

Publication No._____

Xiaoxia Cui, Ph.D.

The University of Texas at Austin, 2005

Supervisor: Alan M. Lambowitz

Group II introns are both catalytic RNAs and retrotransposable elements. Group II intron-encoded proteins (IEPs) have maturase activity, which promotes intron RNA splicing, and reverse transcriptase activity, which functions in intron mobility. Previous studies of the *Lactococcus lactis* Ll.LtrB intron suggested a model in which its IEP binds first to a high-affinity binding site in intron subdomain DIVa and then makes additional contacts with the conserved catalytic core to stabilize the active RNA structure. In the absence of DIVa, the IEP promotes residual splicing by binding directly to the catalytic core. I developed *E. coli* genetic assays to detect *in vivo* splicing of the Ll.LtrB intron and identify regions in the IEP essential for interacting with different parts of the intron RNA. Mutational and biochemical analysis combined with three-dimensional structural modeling support the hypothesis that the extended N-terminal finger region of LtrA is involved in high-affinity binding of DIVa, possibly forming a binding pocket in combination with parts of the thumb domain (domain X), while other regions of the RT and X domains are potentially involved in binding the catalytic core.

The Ll.LtrB intron works very efficiently for gene targeting in bacteria, and it is desirable to target mammalian genomes, for which efficient means of manipulation are lacking. I developed an expression system to produce Ll.LtrB intron RNA and IEP in cultured human cells and found that the expressed intron splices *in vivo*. I also explored different methods for introducing RNPs into cells, including electroporation and injection, and was able to detect a chromosomal targeting event with RNP electroporation. With improvements in targeting efficiency, group II introns would be generally useful for functional genomics and gene therapy.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Group II introns are ribozymes that catalyze their own splicing. Mechanistically, group II introns splice like eukaryotic spliceosomal introns and have been hypothesized to be ancestors of spliceosomes (see section 1.4.3). First identified in 1982 (Davies et al., 1982; Michel et al., 1982; Schmelzer et al., 1982) and widely distributed in bacterial and organellar genomes, some group II introns encode a multifunctional protein and are retrotransposable elements (Lambowitz & Zimmerly, 2004). The intron-encoded protein (IEP) assists the intron RNA in various reactions *in vivo* and *in vitro*, including intron splicing, recognition and insertion of the intron RNA into DNA targets, and reverse transcription of the inserted intron into DNA. The reverse transcriptases (RTs) encoded by mobile group II introns are related to those encoded by non-long terminal repeat (non-LTR) retrotransposons and telomerase. Group II introns provide a model for understanding how proteins interact with catalytic RNAs (see section 1.4.3). In addition, mobile group II introns have potential applications in functional genomics and gene therapy.

1.1 Structural characteristics of group II intron RNAs

Group II introns lack significant primary sequence conservation but have characteristic secondary and tertiary structures (Michel & Ferat, 1995). Subtle variations in secondary and tertiary features were used as criteria to categorize group II introns into subgroups IIA and IIB, each of which can be further divided into two subclasses: A1 and A2; B1 and B2 (Toor et al., 2001). Both IIA and IIB introns include protein-coding introns that are retroelements. More recently, certain bacterial introns were classified as IIC introns. All known IIC introns encode an open reading frame

(http://www.fp.ucalgary.ca/group_2introns/). The signature secondary structure consists of six stem-loop domains denoted DI to DVI (Figure 1.1).

Domains I and V are essential for the catalytic activity. Domain I contains exon-binding sequences (EBS) and other sequence motifs involved in intra- and inter-domain tertiary interactions (Figure1.1) and acts as a scaffold for the folding of the intron RNA (Qin & Pyle, 1998). Domain V is the most phylogenetically conserved portion of the intron and docks onto domain I to form the minimal catalytic core (Michel et al., 1989; Qin & Pyle, 1998).

DII and DIII are not required for catalysis but contribute to optimal splicing. At least in the yeast intron aI5γ, DIII is a catalytic effector during the formation of the catalytic core. DII mediates conformational changes during splicing, including positioning of 5' splice site and recruitment of domain III to the catalytic core (Fedorova et al., 2003). DIV is the least conserved domain and not essential for RNA catalysis, but always encodes the IEP ORF in protein-coding introns. DVI is not conserved in primary sequence but contains structural determinants for branch-site recognition, including the most common feature, a bulged adenosine (Chu et al., 2001).

Remarkably, most group II domains, when transcribed *in vitro* as separate molecules, can fold independently and maintain discrete functions in intron catalysis (Qin & Pyle, 1998). *In vivo*, two group II introns in the green alga *Chlamydomonas reinhardtii* chloroplast are transcribed in pieces and splice in *trans* with the help of proteins (Perron et al., 1999). The ability of group II intron domains to function *in trans* demonstrates that inter-domain tertiary interactions correctly determine how these substructures associate with each other and that group II intron domains could be prototypes for snRNAs in the spliceosome.

1.2 CHARACTERISTICS OF GROUP II INTRON IEPS

Group II intron IEPs are intron-specific protein factors for splicing and/or mobility. Generally, four distinct domains are defined in group II IEPs. A reverse transcriptase (RT) domain is located at the N-terminus, followed by domain X and the C-terminal DNA-binding (D) and endonuclease (En) domains (Figure 1.2). Four enzymatic activities are associated with group II IEPs. The maturase activity refers to the ability of an IEP to bind and induce conformational changes in an intron RNA during splicing (Saldanha et al., 1999; Wank et al., 1999; Matsuura et al., 2001; Noah & Lambowitz, 2003). Reverse transcriptase, DNA binding and DNA endonuclease activities play key roles in mobility. The N-terminal RT domain is related to the RTs of non-LTR retrotransposons, such as the *Bombyx mori* R2Bm element and the human LINE elements (Eickbush, 1994), and carries out the essential step of reverse transcription in intron mobility. The D domain recognizes target DNA, and the En domain cleaves bottom strand of the target site. The En domain contains conserved sequence motifs characteristic of the H-N-H family of DNA endonucleases, interspersed with two pairs of conserved cysteine residues, which are required to maintain the active structure of the domain (San Filippo & Lambowitz, 2002). Some group II IEPs contain all four domains, whereas others only have RT and X domains (http://www.fp.ucalgary.ca/group2introns/).

There are two major hypotheses about the origin of ORF-encoding group II introns. One is that IEP ORFs were recruited by preexisting self-splicing RNA structures (Kennell et al., 1993; Wank, et al., 1999). The other model speculates that group II intron RNA structure evolved from a retroelement RNA (Curcio & Belfort, 1996). Over 50 group II introns have been identified across the prokaryotic spectrum, including Gram-negative, Gram-positive, proteobacteria, cyanobacteria, and archaea (Dai et al., 2003).

Virtually, all bacterial group II introns identified so far encode IEPs. Without identifying the ultimate origin, analyses of bacterial group II introns supported a new model for group II intron evolution, termed the retroelement ancestor hypothesis (Toor, et al., 2001). This hypothesis suggests that modern mobile group II introns may have come from bacterial ORF-containing group II introns, which horizontally spread to mitochondria and chloroplasts. The intron RNAs and IEPs then co-evolved in both bacteria and organelles. Toor et al. (2001) also presented evidence that most ORF-less introns are derivatives of ORF-containing introns. In higher organisms, organellar group II introns lost or have mutated ORFs and rely on cellular factors for splicing (Toor et al., 2001).

1.3 OVERVIEW OF STUDIES ON GROUP II INTRON SPLICING

Group II intron RNAs are ribozymes, some of which are capable of splicing *in vitro* in the absence of protein (Michel et al., 1989). However, self-splicing requuires extreme conditions including high salt and magnesium concentrations. The presence of the maturase protein lowers the required concentrations for both salt and magnesium, and protein-dependent splicing is more rapid than self-splicing (Saldanha et al., 1999). The protein helps the intron RNA to fold and maintains active structures that are not optimally formed with high salt and magnesium (Saldanha et al., 1999).

1.3.1 Splicing mechanism

Group II intron splicing occurs via two transesterification reactions. First, the bulged A residue in DVI attacks the first nucleotide of the intron at the 5' splice site to form a 2', 5'-phospodiester bond and to release the 5' exon. The 3' hydroxyl of the 5' exon then attacks the phosphodiester bond at the 3' splice site, resulting in ligated exons and an intron lariat (Figure 1.3a). The first step of splicing is reversible and rate limiting.

This reversibility may provide a proofreading mechanism for the 5' splice site cleavage (Chin & Pyle, 1995). After protein-dependent splicing, the IEP stays bound to the excised lariat RNA, forming ribonucleoprotein (RNP) particles, which are functional entities in mobility.

The first step of splicing can also occur through hydrolysis, in which a water molecule attacks the phosphodiester bond at the 5' splice site (Figure 1.3b). After an essentially identical second step, a linear intron, instead of a lariat intron, is released. The hydrolysis pathway is slow but irreversible and is favored by potassium chloride in the reaction *in vitro* (Chin & Pyle, 1995). *In vivo*, this pathway is observed in yeast mitochondrial intron aI5γ with the bulged A residue mutated (Podar et al., 1998) and in a barley chloroplast intron lacking a bulged A in domain VI (Vogel & Borner, 2002).

1.3.2 The *Lactococcus lactis* group II intron Ll.LtrB

The *Lactococcus lactis* group II intron Ll.LtrB belongs to the IIA1 subgroup and is by far the best biochemically studied RT/maturase-coding group II intron, mainly because it expresses and splices efficiently in *E. coli* (Matsuura et al., 1997). The Ll.LtrB IEP, LtrA, has RT, maturase, and DNA endonuclease activities (Matsuura et al., 1997). Biochemical analysis showed that LtrA is an intron-specific splicing factor that binds the intron RNA with high affinity (kinetically measured apparent $K_d \leq 0.12$ pM) and is by itself sufficient for maturase activity (Saldanha et al., 1999). A series of deletions of the intron RNA were analyzed and showed that a high-affinity binding site for LtrA is located in DIVa (Wank et al., 1999). The high-affinity binding site was further localized by *in vitro* selection and mutagenesis to the distal end of the DIVa, including the ribosome-binding site (RBS) and ATG codon for LtrA translation. The binding of LtrA to the RBS was also shown to down-regulate its own translation (Singh et al., 2002). Matsuura et al. (2001) identified potential low-affinity protein-binding sites in the

conserved catalytic core and demonstrated that binding of LtrA induced the intron RNA to form critical tertiary interactions for splicing. UV-cross linking experiments showed that LtrA also stabilizes specific tertiary interactions. Magnesium, on the other hand, stabilizes all tertiary interactions including non-native ones, and may slow down the splicing reaction because the misfolded RNAs need time to refold into active structures (Noah & Lambowitz, 2003).

Modifying the intron RNA has been a major approach for studying the RNA/protein interactions in the Ll.LtrB intron splicing. It is also important to address the question of how the protein interacts with the RNA, which was one main goal of my dissertation research.

1.3.3 Other splicing factors

In addition to maturases, a number of other proteins have been found to be involved in group II intron splicing. These splicing factors are summarized here in two categories: maturase-related and maturase-unrelated.

The maturase-related category includes MatK proteins and nMat proteins. MatK proteins are encoded in a group II intron in the $tRNA^{lys}$ genes of higher plant chloroplasts or by freestanding ORFs in the residual plastid genomes of non-photosynthetic plants (Mohr et al., 1993; Vogel et al., 1997; Young & dePamphilis, 2000).). MatK ORFs contain an X domain and a degenerate RT domain, but lack the DNA-binding and DNA-endonuclease domains. They may function in splicing multiple chloroplast group II introns, many of which do not encode ORFs themselves (Mohr et al., 1993; Vogel et al., 1997; Young & dePamphilis, 2000).

nMat proteins were identified in higher plant nuclear genomes by sequence comparisons and contain two families, each with two members: nMat-1a, nMat-1b, nMat-2a, and nMat-2b. Both protein families have characteristic group II intron IEP RT and X

domains. The nMat2 proteins also have D and En domains, and the nMat1 proteins have a novel C-terminal domain. All four proteins have a conserved X domain but mutated RT (nMat2) or D/En domains (nMat1), implying that they may retain the ability to function in group II intron splicing but are unable to fully support intron mobility. All nMat proteins are encoded by nuclear genes with no association with group II intron structures. The proteins have mt localization sequences and may be imported into that organelle to facilitate the splicing of ORF-less group II introns (Mohr & Lambowitz, 2003).

Besides maturase-related proteins, several other group II intron splicing factors have been identified. Two of these proteins were identified in yeast mitochondria. The Mrs2p protein seemed to be the only splicing factor essential for splicing of all group II introns in the yeast mitochondria. However, Mrs2p is a magnesium transporter, and facilitates splicing indirectly by increasing the Mg^{2+} concentration (Gregan et al., 2001). Mss116p is an RNA helicase encoded by a nuclear gene and is transported to the mitochondrion. Mss116p is required for the efficient splicing of all group I and II introns and acts as an RNA chaperone likely to disrupt misfolded RNA structures (Huang et al., 2004). In addition to participating in the intron splicing in the yeast mitochondrion, Mss116p also functions in translation (Seraphin et al., 1989).

The CRS1 and CRS2 proteins from maize were identified as splicing factors via a genetic screening. CRS1 helps splice a subgroup IIA intron, atpF, whereas CRS2 functions in splicing of multiple subgroup IIB introns (Till et al., 2001). Both proteins function in the context of large RNPs containing intron RNAs (Jenkins & Barkan, 2001; Till et al., 2001). In addition to helping intron splicing, CRS1 may play a role in the chloroplast translation, and CRS2 is a homolog of peptidyl-tRNA hydrolase enzymes (Till et al., 2001). Additionally, two more splicing factors were shown to interact with CRS2 through a yeast two hybrid screen: CRS2-associated factors 1 and 2 (CAF1 and

CAF2). Each is present in RNPs containing CRS2 and a distinct set of intron RNAs. In rare cases, some introns require both CAF proteins for splicing. Like CRS1, CAF1 and CAF2 contain an RNA-binding domain of a novel family named the Chloroplast RNA splicing and ribosome maturation (CRM) domains (Ostheimer et al., 2003).

Three Raa proteins were found to function *in trans* splicing of different group II introns in algae chloroplasts (Perron et al., 1999; Rivier et al., 2001; Perron et al., 2004). Like the CRS proteins, Raa proteins are also associated with large intron-containing RNPs (Perron et al., 2004). The Raa2 protein is related to pseudouridine synthases, but the pseudouridine synthetase activity is not required for the *trans*-splicing reaction (Perron et al., 1999).

Most recently, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) has been found to help splice yeast bI1 intron *in vitro*. However, since GAPDH does not enter mitochondria, it is unlikely that the same interaction happens *in vivo* (Bock-Taferner & Wank, 2004).

1.3.4 Evolutionary aspects

Two classes of spliceosomal introns, denoted U2-dependent and U12-dependent, have been identified in eukaryotes. Each class is spliced by a different spliceosome assembled from five small nuclear RNAs and numerous proteins (Zhu & Brendel, 2003). Group II introns are hypothesized to be ancestors of spliceosomes (Michel & Ferat, 1995). The following evidence supports the hypothesis. First, group II introns and spliceosomes use two essentially identical transesterification reactions for splicing (Guthrie, 1991; Steiz & Steiz, 1993; Sontheimer et al., 1999; Gordon et al., 2000). Spliceosomal introns could participate in alternative splicing by selecting different splicing sites. Interestingly, a bacterial group II intron was also found to have a primitive form of alternative splicing (Robart et al., 2004). Second, as in group II intron splicing,

RNA is believed to be the catalytic unit in spliceosomes (Will & Luhrmann, 2001; Villa et al., 2002; Valadkhan & Manley, 2003). Third, structural similarities were identified between group II intron domains and snRNAs (Yu et al., 1995; Sashital et al., 2004). Segments of a group II intron and certain snRNAs can replace each other for function. For examples, a group II intron DV can function in the U12-dependent spliceosome *in vivo* in place of the U6atac snRNA stem loop (Shukla & Padgett, 2002), and the U5 snRNA *trans*-activates group II intron splicing by replacing its ID3 structure (Hetzer et al., 1997). If group II introns are in fact ancestors of spliceosomes, the much simpler group II intron system provides a unique platform for studying the splicing reactions.

1.4 OVERVIEW OF STUDIES ON GROUP II INTRON MOBILITY

Retroelements propagate via RNA intermediates and reverse transcription. This broad category includes, but is not limited to, LTR and non-LTR retrotransposons, retrovirus, telomerase, and mobile group II introns (Kazazian, 2004). Group II introns and non-LTR retrotransposons use the similar mechanism to integrate into DNA targets, and RTs encoded by these two groups are phylogenetically related (Eickbush, 1994).

1.4.1 General mobility mechanism—target DNA-primed reverse transcription

Non-LTR retrotransposons use a pathway termed **t**arget DNA-**p**rimed **r**everse **t**ranscription (TPRT) during integration, where the 3' end of a nicked genomic DNA strand is used as a primer for reverse transcription of the element's RNA sequences (Figure 1.4, Eickbush, 1992). Many non-LTR retrotransposons insert nonspecifically at random nicks, whereas some generate specific cleavages. For example, the R2Bm element from the silkworm, *Bombyx mori*, inserts into a specific site in the 28S rRNA gene (Luan et al., 1993). However, TPRT only accomplishes the 3' end insertion of R2Bm, and the host homologous recombination is required for 5'-end insertion (Fujimoto

et al., 2004). Human LINE 1 elements appear to integrate nonspecifically into the genome likely through TRPT (Moran et al., 1996; Feng et al., 1996; Morrish et al., 2002). In addition to non-LTR retrotransposons, telomerase also uses TPRT to synthesize the highly specific telomere repeat sequences at most eukaryotic chromosomal termini (Lingner et al., 1997).

1.4.2 Overview of the work on yeast intron mobility

Mobile group II introns are site-specific retroelements and insert at high frequencies into intronless alleles, a process termed retrohoming, and at low frequencies, to ectopic sites that resemble the natural homing site, a process termed retrotransposition (Lambowitz & Zimmerly, 2004). Pioneering work on group II intron mobility was done mainly with yeast introns aI1 and aI2, which are the first and second introns of the *COX1* gene encoding subunit 1 of cytochrome oxidase. Zimmerly et al. (1995a, b) showed that TPRT was the mechanism for aI1 and aI2 homing, and that the intron RNA cleaves the top strand. Homing is mediated by RNP particles, complexes assembled from the intron lariat and the maturase protein. The intron RNA in an RNP particle inserts into the sense strand of the target DNA, and the maturase protein cleaves the bottom strand after the 10^{th} nucleotide downstream of the insertion site, and uses the liberated 3'-OH to prime cDNA synthesis. In the major retrohoming pathway, the intron cDNA invades an intronless allele to initiate reverse splicing and reverse transcription stimulate repair/recombination event that leads to intron insertion with subsequent coconversion of upstream exon sequences. Retrohoming can also be completed by complete reverse splicing, synthesis of a full-length of cDNA, and repairing the gap in the top strand, independent of homologous recombination. Finally, retrohoming can occur independent of RT activity after target-site cleavage, likely through double strand break repair recombination and characterized by bidirectional coconversion of flanking exon

squences. The DNA target site, intron, and the environment that retrohoming takes place are contributing factors to which pathway is chosen (Eskes et al., 2000).

During target recognition, EBS and δ sequences in the intron RNA base pair with IBS and δ' sequences in the DNA target site. Modifying these base-pairing interactions changes the targeting specificity (Eskes et al., 1997; Guo et al., 1997), raising the possibility of targeting group II introns to desired sites. Unfortunately, neither aI1 nor aI2-encoded proteins have been expressed and correctly processed other than in the yeast mitochondria. This largely limits the usefulness of these introns in targeting elsewhere.

1.4.3 Studies on the *Lactococcus lactis* LtrB intron

The ability to express the Ll.LtrB intron in *E. coli* made it possible to study the Ll.LtrB intron mobility in great detail. Mechanistically, the intron was shown to use TPRT for mobility via complete reverse splicing and reverse transcription, and modifying EBS and δ sequences of the intron changes the target site specificity for reverse splicing (Matsuura et al., 1997). Unlike the situation for yeast introns, the Ll.LtrB intron cannot use an RT-independent pathway, in which after reverse transcription, the intron cDNA integrates by a DNA repair step, dependent on homologous recombination (Cousineau et al., 1998). During mobility, the RNPs bind DNA non-specifically and search for target sites (Aizawa et al., 2003). At a target site, the IEP first recognizes a few bases in the distal 5' exon region via major groove and phosphate-backbone interactions, which triggers the unwinding of the double-strand target DNA, enabling base pairing interactions between the intron RNA and target DNA for reverse splicing of the intron in the top strand (Singh & Lambowitz, 2001). Bottom-strand cleavage then occurs with additional interactions between the IEP and the 3' exon (San Filippo & Lambowitz, 2002). However, top-strand insertion is not required for bottom-strand cleavage (Aizawa et al., 2003). After reverse transcription, the host RNase H activity probably digests the

intron RNA inserted in the top strand, and the host repair machinery synthesizes the second strand. Interestingly, when the RT is defective in cleaving the bottom strand, reverse transcription proceeds using a nascent strand at a DNA replication fork as the primer, and mobility still can be completed at a lower frequency (Zhong & Lambowitz, 2003). This could be one of the pathways for the mobility of introns encoding endonuclease-less IEPs (Munoz-Adelantado et al., 2003), as well as a mechanism for group II intron retrotransposition, which is independent of the D and En domains (Ichiyanagi et al., 2002).

Using group II introns to target desired genes has always been one of the most favorable prospects of group II intron studies. Remarkable progress has been made in bacteria in using group II introns as tools for gene disruption. Target-site recognition rules were obtained (Mohr et al., 2000; Guo et al., 2000; Zhong et al., 2003) and used to develop an algorithm that designs introns to target a given gene with high efficiency (Perutka et al., 2004). Genomes of different bacteria were targeted by group II introns (Karberg et al., 2001; Zhong et al., 2002), and in *L. lactis*, conditional knockouts using group II introns were obtained (Frazier et al., 2003).

1.4.4 Significance of intron mobility

Using group II introns in higher organisms is of great interest because of the difficulties of manipulating their more complex genomes. Guo et al. (2000) presented initial evidence that group II RNPs were active in human cells. The second part of my dissertation focuses on using group II introns in mammalian targeting.

In addition to their practical potentials, group II introns may be ancestors of telomerase and non-LTR retroelements, such as the human LINE elements, which use a similar TPRT mechanism for mobility (Lambowitz & Zimmerly, 2004). Studies of group II intron mobility may also advance the understanding of these retroelements.

1.5 OVERVIEW OF THE CHAPTERS

This dissertation research has two main themes: maturase/intron RNA interactions during group II intron splicing and utilizing group II introns to target mammalian genomes. Two chapters are devoted to each of these areas, followed by a chapter describing Materials and Methods.

Chapter 2 focuses on the identification of LtrA domains that function in the Ll.LtrB intron splicing. I will describe a genetic system I developed to monitor *in vivo* splicing of the Ll.LtrB intron in *E. coli*. Based on this system, I constructed a random LtrA library and screened for splicing-competent mutants. Statistical analysis of the ratios of missense to silent mutations in different regions demonstrated that both RT and X domains are required for splicing. Biochemical analysis suggested that the N-terminus of the RT domain is involved in high-affinity binding of the intron RNA, while domain X participates in interaction with the intron catalytic core. Chapter 3 is a continuation of the studies of chapter 2, where I did another genetic screening in the absence of the high-affinity protein-binding site of the intron. By comparing the two sets of data, several regions were identified as candidates for directly interacting with the DIVa sequence. A series of site-directed mutations were tested to further pinpoint the interaction spots. The mutagenesis data were interpreted by using a newly developed LtrA structural model based on HIV-1 RT structures. This part of the research provides a detailed model of how LtrA binds DIVa. Differences between *in vivo* and *in vitro* data are discussed. Chapter 4 describes work on expressing the Ll.LtrB intron RNA and the LtrA protein in human cells. Major findings of this part include that maturase-promoted splicing can take place in human cells with expressed components and that some protein-independent splicing can occur when high magnesium is present in the medium. The expressed LtrA protein had RT activity and was localized to the nucleus. In chapter 5, RNP particles purified

from *E. coli* were used to target human cells directly. A series of *in vitro* experiments was done to show that RNPs are capable of identifying target sites in the context of the human genome and in nuclear lysates. Electroporation could deliver RNPs into the cells without compromising RNP activity. A chromosomal rRNA gene in human cells was targeted with electroporated RNPs. Thus, using RNPs for targeting is feasible but the efficiencies need to be increased. Chapter 6 describes the methods used in this research.

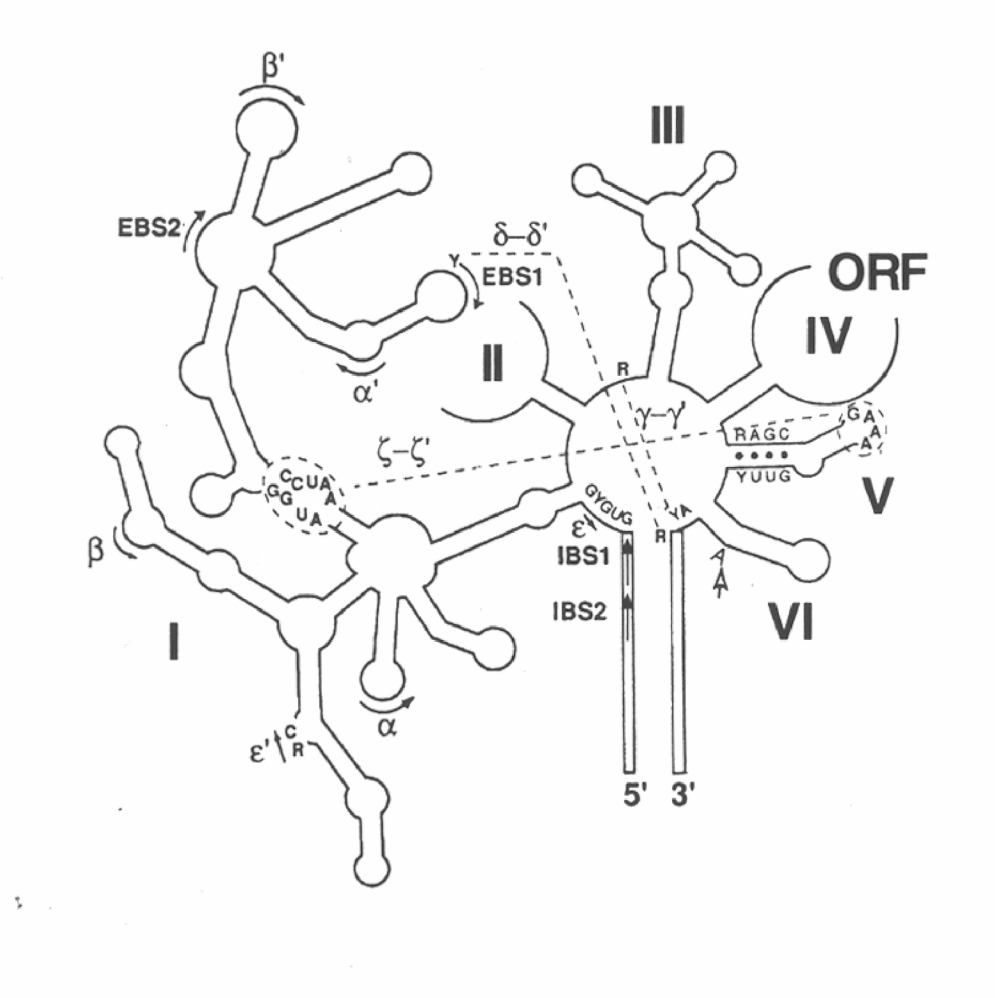


Figure 1.1 Characteristic secondary structure and predicted tertiary interactions of group II introns.

The six domains of a group IIA intron are labeled as I to VI surrounding the central wheel. IBS1 and 2 sequences in the 5' exon and EBS1 and 2 sequences in the intron are labeled. Tertiary interactions are indicated as pairs in Greek letters.

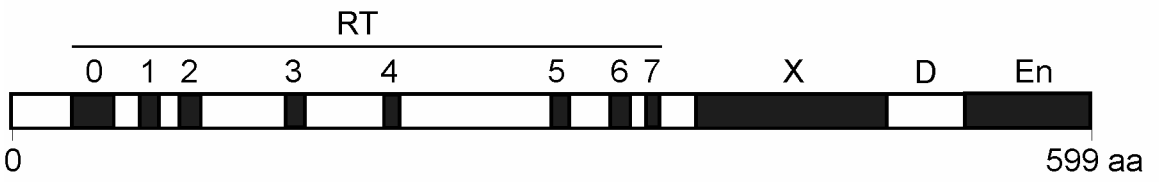


Figure 1.2 Schematic of the Ll.LtrB intron-encoded protein, LtrA, showing general features of group II intron IEPs.

Four domains are present. The RT domain is at the N-terminus of the protein, with seven conservative motifs, RT1 to 7, found in the fingers and palm of retroviral RTs, and RT0, a motif characteristic of non-LTR RTs (Eickbush, 1994). Domain X, believed to be involved in splicing and historically called the maturase domain (Mohr et al., 1993), follows the RT domain and functions in RNA splicing. Further downstream are the DNA-binding (D) and DNA endonuclease (En) domains, which are essential for mobility.

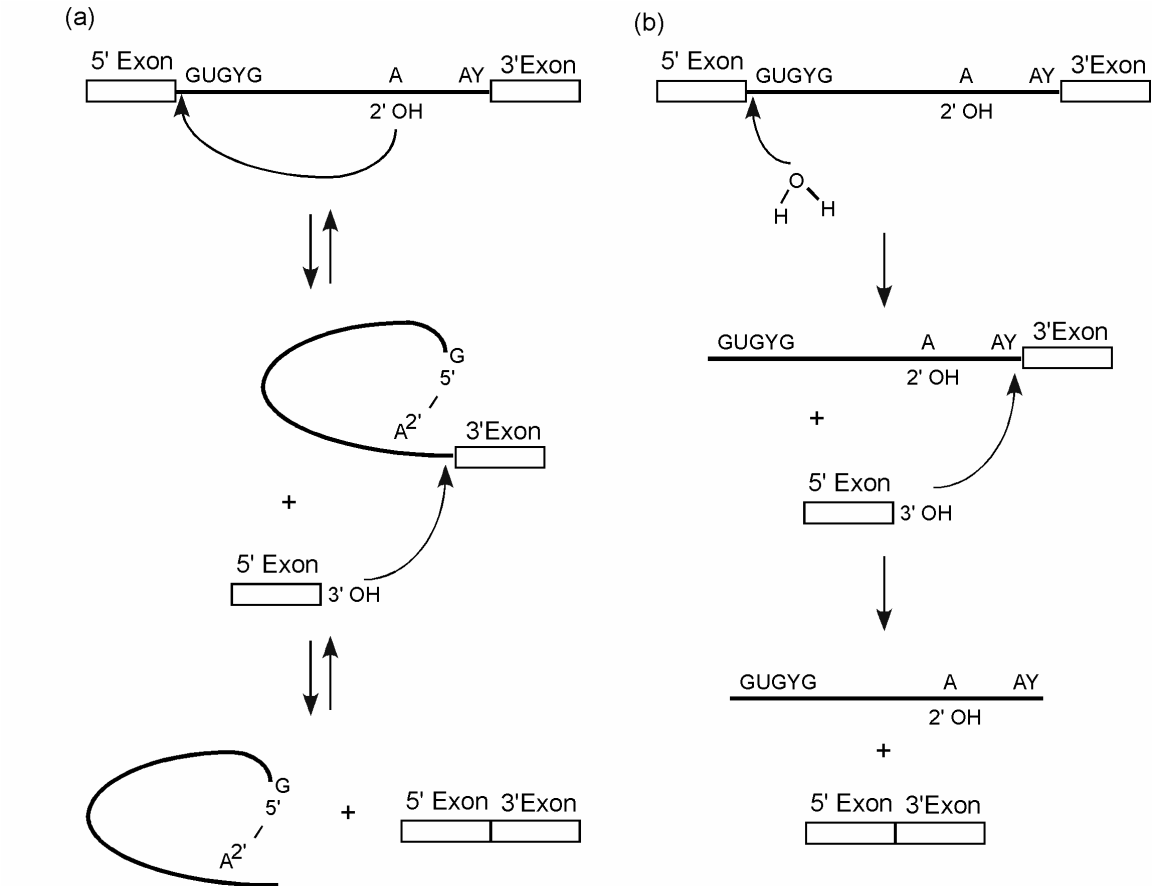


Figure 1.3 Splicing mechanisms of group II introns.

(a) The branching pathway. In this pathway, the 2' hydroxyl group of the bulged A in domain VI attacks the 5' splice site and form a 2', 5' phospodiester bond, releasing the 5' exon. The 3' hydroxyl group of the 5' exon then attacks the 3' splice site and generates splicing products: lariat intron and ligated exons. The first transesterification step is reversible and rate limiting, and the second step is rapid and drives the reaction to completion (Qin & Pyle, 1998). (b) The hydrolysis pathway. In this pathway, instead of the bulged A, a water molecule initiates the reaction by attacking the 5' splice site to free the 5' exon irreversibly. After an essentially identical second step, ligated exons and linear intron are formed (Chin & Pyle, 1995).

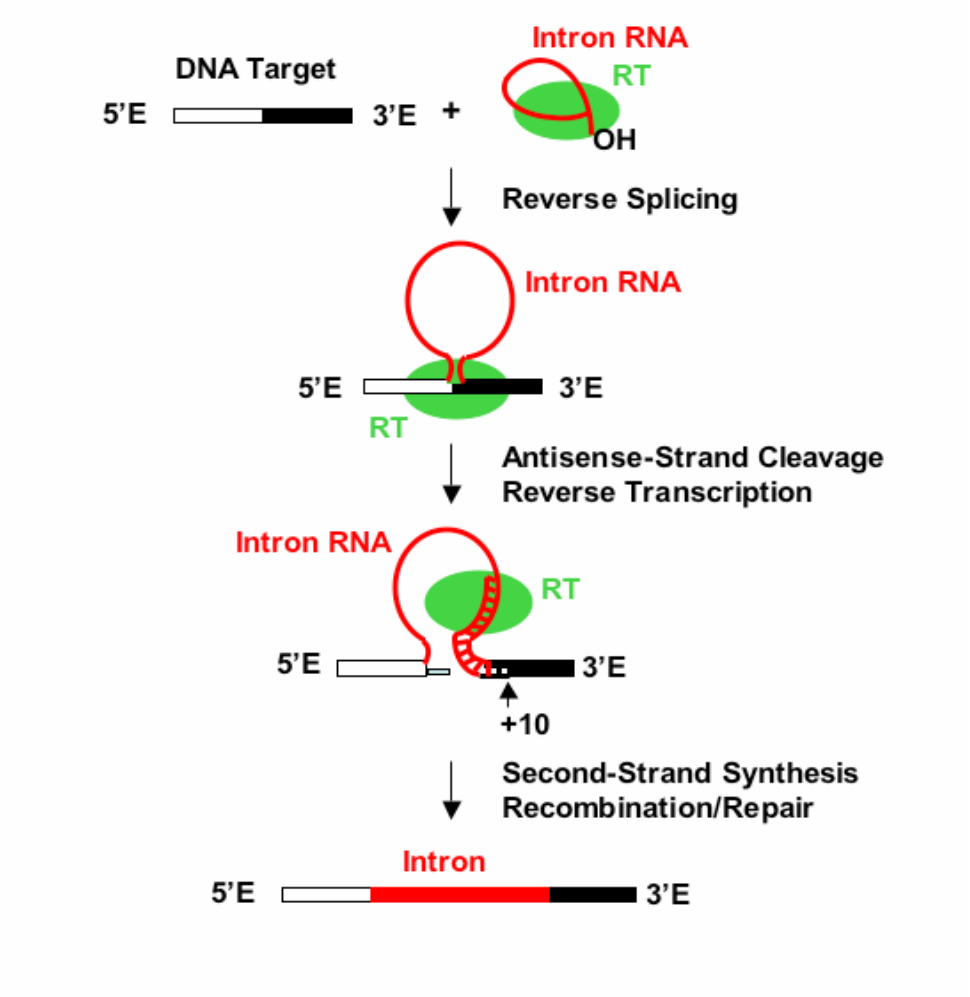


Figure 1.4 Schematic of the TPRT mechanism used by group II introns to insert into DNA target sites.

An RNP particle consisting of the lariat intron and the IEP recognizes the target sequence, and the intron RNA reverse splices into the sense strand of the insertion site. The IEP uses its endonuclease activity to cleave the antisense-strand and its RT activity to reverse transcribe the intron RNA using the 3' end of the cleaved antisense DNA as the primer. The second strand is then synthesized and gaps repaired by host factors (Zimmerly et al., 1995a, b; Yang et al., 1996; Eskes et al., 1997).

Chapter 2: An *E. coli* Genetic Assay for Maturase-promoted Group II Intron Splicing and Delineation of Regions of the IEP Required for Splicing

The best characterized group II intron IEPs are encoded by the yeast mitochondrial introns aI1 and aI2 and the *Lactococcus lactis* intron Ll.LtrB. These proteins contain all four conserved domains (see section 1.2), RT, X, DNA-binding (D) and DNA endonuclease (En) domains. The N-terminal RT domain contains conserved amino acid sequence blocks 1-7 (RT-1 to -7) characteristic of the fingers and palm of retroviral RTs, preceded by an upstream region, RT-0 (previously denoted Z) found in the RTs of non-LTR-retrotransposons, such as human LINE elements and insect R2 element (Xiong & Eickbush, 1990; Zimmerly et al., 2001). The RT domain is followed by domain X, which was identified as a site of mutations affecting maturase activity (Mohr et al., 1993; Moran et al., 1994). Domain X is located in the position corresponding to the thumb and connection domains of retroviral RTs (Mohr et al., 1993). Finally, the C-terminal D and En domains function in intron mobility. The En domain contains conserved amino acid sequence motifs characteristic of the H-N-H family of DNA endonucleases, interspersed with two pairs of conserved cysteine residues, which are thought to help maintain the active structure of the domain (San Filippo & Lambowitz, 2002).

In addition to being the site of mutations that affect RNA splicing, the association of domain X with maturase activity is supported by the pattern of conserved and nonconserved regions in degenerate group II IEPs, which have lost mobility functions but may retain RNA splicing activity (Mohr et al., 1993; Mohr & Lambowitz, 2003). These include two classes of maturase-related splicing factors: higher plant chloroplast MatK proteins and higher plant mitochondrial nMat proteins (see section 1.3.1). MatK proteins

lack the En domain and have degenerate RT domains in which only RT-5 to -7 are recognizable, but still retain a conserved domain X (Mohr et al., 2003; Vogel et al., 1997; Young & dePamphilis, 2000). All of the nMat proteins have a putative mitochondrial targeting sequence and contain RT and X domains, and nMat-2 proteins also contain cognates of the D and En domains. However, only domain X is conserved in all cases, while the RT and En domains generally have mutations, suggesting loss of mobility functions, similar to the situation for MatK proteins. MatK and nMat proteins may function in the splicing of multiple chloroplast and mitochondrial group II introns, respectively, most of which do not themselves encode ORFs (Wolfe et al., 1991; Vogel et al., 1999; Mohr & Lambowitz, 2003).

To study maturases biochemically, an *E. coli* expression system for the *L. lactis* Ll.LtrB intron was developed, which allows purification of large quantities of the IEP, denoted LtrA protein (Saldanha et al., 1999; Matsuura et al., 1997). Studies using the purified protein showed that LtrA is an intron-specific splicing factor, which binds tightly and specifically to the Ll.LtrB intron but not other group II introns, and is by itself sufficient to promote the splicing of the Ll.LtrB intron *in vitro* at a low Mg^{2+} concentration (5 mM), where the intron cannot self-splice efficiently. LtrA binds the intron RNA as a dimer, the same active form as the HIV-1 RT (Saldanha et al., 1999). After splicing, the IEP remains bound to the excised intron lariat RNA and promotes intron mobility by target-DNA primed reverse transcription (TPRT; see section 1.5.1). Thus, the interaction of the maturase with the intron RNA is critical for both RNA splicing and intron mobility.

Biochemical studies also identified DIVa, an idiosyncratic structure that contains the ribosome-binding site and initiation codon of the LtrA ORF, as the primary high-affinity binding site of LtrA (Wank et al., 1999; Fig. 2.1). The binding of LtrA to DIVa

is thought to occur first and occludes the ribosome-binding site, thereby down-regulating translation and preventing ribosome entry into the intron, which might otherwise impede splicing (Singh et al., 2002). After binding to DIVa, the LtrA protein makes additional contacts with conserved catalytic core regions, including DI and possibly parts of DII and DVI, to stabilize RNA tertiary structure required for ribozyme activity (Matsuura et al., 2001; Noah & Lambowitz, 2003).

Notably, when DIVa is deleted, the LtrA protein can still promote residual splicing *in vitro* by binding directly to the catalytic core. However, the stability of the complex is decreased by ~10^5, and the residual splicing is slower even at saturating protein concentrations, suggesting that DIVa also facilitates one or more rate-limiting conformational changes to the active RNA structure (Wank et al., 1999; Matsuura et al., 2001). Recent UV-cross-linking experiments indicated that the catalytically active RNA structure might form at least transiently at low Mg^{2+} concentration, raising the possibility that after binding to DIVa, the maturase functions at least in part by tertiary structure capture (Noah et al., 2003). The finding that the IEP interacts with conserved as well as idiosyncratic regions implies that it contains multiple binding sites that interact with different regions of the intron RNA. It also suggests how maturases could have evolved from intron-specific into a general group II intron splicing factor, as may have occurred for the plant MatK and nMat proteins.

Here, to further study RNA-protein interactions in group II intron splicing, I developed an *E. coli* genetic assay for maturase-dependent splicing by linking the splicing of the Ll.LtrB intron to the expression of GFP. I then used this assay to analyze the *in vivo* splicing of wild-type and mutant introns and to delineate regions of the maturase required for RNA splicing. My results indicate that the IEP functions in splicing by using the RT and X domains to bind different regions of the intron RNA and suggest

that the IEP may have adapted to function in splicing, at least in part, via interactions used to recognize the intron RNA as a template for reverse transcription.

2.1 AN *E. COLI* GENETIC ASSAY FOR MATURASE-PROMOTED SPLICING OF THE LL.LTRB INTRON

I developed an *E. coli* genetic assay for maturase activity by linking the splicing of the Ll.LtrB intron to the expression of green fluorescent protein (GFP). Insertion of the intron RNA into reporter genes is not trivial because group II intron splicing requires base pairing of intron sequences EBS1, EBS2 and δ located in DI to 5'- and 3'-exon sequences IBS1, IBS2, and δ' (Figure 2.1; Michel & Ferat, 1995). EBS and IBS denoted exon- and intron-binding sites, respectively. I solved this problem by inserting the intron and minimum-sized flanking exon sequences from the *ltrB* gene at the N-terminus of the GFP ORF, with the 5' exon fused in-frame to a short ϕ10 leader sequence, and the 3' exon fused to codon 2 of GFP (Figure 2.2). Splicing then leads to the synthesis of GFP with a short N-terminal extension (42 amino acid residues) corresponding to the ϕ10 leader and ligated exons, whereas in the absence of splicing, translation is terminated by multiple stop codons within the intron (Figure 2.2(c)). The wild-type construct, pALG1-WT, contains the full-length Ll.LtrB intron with the LtrA protein encoded in DIV (Figure 2.2(a)), while pALG2-WT contains a 0.9-kb Ll.LtrB-ΔORF intron with the LtrA protein expressed from a position just downstream of the 3'-exon/GFP sequence (referred to as the downstream *cis* position; Figure 2.2(b)).

To assay splicing, plasmids containing the *ltrB*/GFP fusion were transformed into *E. coli* HMS174(DE3), which contains an integrated T7 RNA polymerase gene under the control of the *lac*UV5 promoter. After induction with IPTG, single-cell fluorescence was measured with a fluorescence-activated cell sorter (FACS). As shown in Figure 2.3(a), cells transformed with pALG1-WT and induced with IPTG gave a peak of 820

fluorescence units, whereas a similar fluorescence increase was not seen without IPTG induction, nor for cells transformed with the vector pAT (Figure 2.3(b) and (c), respectively). Likewise, no fluorescence increase was seen after IPTG-induction for the splicing-defective mutants pALG1-ΔDV, which is deleted for the catalytically essential intron DV, nor for pALG1-ΔORF, which has a large deletion in the LtrA ORF (Figures 2.3(d) and (e), respectively).

Notably, the splicing of the ΔORF intron measured by GFP expression occurred at wild-type levels in pALG2-WT, in which the LtrA protein is expressed from the downstream *cis* position (peak of 843 units; Figure 2.3(f)), but at a much lower level when LtrA is expressed in *trans* from a separate plasmid (pALG1-ΔORF + LtrA; peak of 48 units; Figure 2.3(g)). Immunoblots of total cellular protein showed that all three constructs produced similar amounts of LtrA (Figure 2.4(a)). Since the plasmid used to express LtrA in *trans* is of higher copy number than that used to express LtrA in *cis*, the similar levels of LtrA suggest that excess, unbound protein is degraded (see below). In addition, when expressed in *trans*, the proportion of soluble LtrA was usually decreased substantially, as shown in Figure 2.4(b) (compare lanes 2 and 5 with lane 6). The *cis* constructs pALG2-ΔDIVa and pALG2-ΔIVa/b, which lack the high-affinity binding site in DIVa, also generally gave a high proportion of insoluble LtrA (Figure 2.4, lanes 7 and 8), while the construct pALG1-ΔDV, which is missing a key region of the intron's catalytic core, gave ~50% lower amounts of soluble LtrA in some experiments, but not others (Figure 2.4, lanes 5, and data not shown). These results are consistent with previous findings that the yield of LtrA was decreased threefold and the protein was largely insoluble when expressed at 37˚C in the absence of the intron RNA (Matsuura et al., 1997), and with *in vitro* experiments demonstrating that LtrA is prone to rapid and irreversible denaturation when incubated by itself in splicing reaction medium at 37˚C

(Saldanha et al., 1999). Together, these findings suggest that the decreased splicing efficiency when LtrA is expressed in *trans* reflects in part that the protein is unstable when it cannot bind rapidly and tightly to the intron RNA.

2.2 RT-PCR AND POISONED PRIMER ASSAYS OF *IN VIVO* SPLICING

The results of the FACS assay were confirmed by using both RT-PCR and poisoned primer extension assays. The RT-PCR assays were carried out with 5'- and 3'-exon primers PHIII and PPI, yielding a 145-bp product for ligated exons and larger products for precursor RNAs (Figure 2.5). In agreement with the FACS assays, the wild-type constructs pALG1-WT, containing the full-length intron, and pALG2-WT, containing the ΔORF intron with LtrA expressed from the downstream *cis* position, gave similar amounts of ligated-exon product (lanes 4 and 7), while the construct in which LtrA is expressed in *trans* from a separate plasmid gave significantly less of this product and showed increased amounts of unspliced precursor RNA (lane 8). The mutants pALG1-ΔORF and pALG1-ΔDV gave no detectable ligated exons (lanes 5 and 6, respectively), even when the PCR was extended to 40 cycles (not shown). I note that the larger RT-PCR product, corresponding to the unspliced precursor RNA, was at most barely detectable for constructs containing the full-length Ll.LtrB intron (pALG1-WT and pALG1-ΔDV), but much stronger with constructs containing the shorter ΔORF-version of the intron. This situation likely reflects both decreased efficiency of RT-PCR with longer templates and greater susceptibility of the larger precursor RNA to nucleolytic degradation (Matsuura et al., 1997; Guo et al., 2000).

In the poisoned primer extension assay (Figure 2.6), the 5'-labeled primer PRET, complementary to a sequence just downstream of the 3'-splice site, was extended with M-MLV RT in the presence of ddCTP to terminate cDNA synthesis at the first G-residue in the template. The resulting products for the ligated exons and unspliced precursor

RNA have lengths of 36 and 38 nts, respectively, with the E1-E2 product resolved as a doublet even when synthesized from an *in vitro* transcript control (lane 2). In agreement with the GFP and RT-PCR assays, the poisoned primer extension assay showed that wild-type construct pALG1-WT and the *cis*-construct pALG2-WT spliced efficiently (lanes 4 and 7; ratio of ligated exons:unspliced precursor = 1.2 ± 0.3 and 1.1 ± 0.07, respectively, where the values are the mean ± standard deviation for three determinations). By contrast, splicing was again undetectable for the Ll.LtrB-ΔORF or ΔDV mutants (ratios < 0.01; lanes 5 and 6), confirming that both the LtrA protein and the ribozyme activity of the intron RNA are required for the Ll.LtrB intron to splice significantly in *E. coli*. When LtrA was expressed in *trans*, the assays showed decreased amounts of both ligated exons and unspliced precursor RNA (lane 8), suggesting that more rapid turnover of unbound precursor RNA also contributes to the decreased splicing efficiency in *trans*.

2.3 SPLICING OF MUTANTS WITH DELETIONS IN DIFFERENT SEGMENTS OF THE INTRON RNA

I next tested the *in vivo* splicing of intron deletion constructs, which had been assayed previously for protein-dependent splicing *in vitro* (Figure 2.1; Wank et al., 1999). The deletions were analyzed in the pALG2 background, in which LtrA is expressed from the downstream *cis* position, where modifications can be introduced into DIV without affecting the LtrA coding sequence. Data for the FACS, RT-PCR, and poisoned primer assays are shown in Figures 2.3, 2.5, and 2.6, respectively. In agreement with previous *in vitro* assays (Wank et al., 1999), constructs deleted for DI or III abolished maturase-promoted splicing *in vivo*, as monitored by all three assays. Surprisingly, a construct deleted for DII showed very weak splicing in the RT-PCR assay with an increased number of cycles (Figure 2.5(b) and (c), lanes 10 and 17), and this splicing was confirmed by sequencing the RT-PCR product.

I also tested the effect of deleting different regions of DIV, which plays a key role in binding the LtrA protein. The deletion of DIVa, which contains the high-affinity LtrA-binding site, alone or together with DIVb, inhibited splicing by ≥ 90%, but not completely (ratios of ligated exons to unspliced precursor RNA in poison primer assay were 0.067 ± 0.006 for ΔDIVa and 0.11 ± 0.04 for ΔDIVa/b, corresponding to 6 and 10%, respectively, of that for the parent construct pALG2; 1.1 ± 0.07; see above). In a construct containing DIVa, deletion of the peripheral regions of DIVb (ΔDIVb1/b2) inhibited splicing by ~30%, while the complete deletion of DIVb (ΔDIVb) inhibited splicing by ~60% (ratio of ligated exons to unspliced precursor RNA 0.67 ± 0.16 and 0.37 ± 0.06, respectively). We confirmed by sequencing that the residual splicing of the ΔDIVa construct yields the correct ligated-exon product and that the residual splicing of all four constructs is LtrA-dependent (not shown). The deletion of all of DIV (pALG2-ΔDIV) reduced splicing to 2-4% of wild type. In pALG2-ΔDIVstem, the DIVstem is deleted, whereas the rest of DIV is intact (Figure 2.1(b)). This mutant spliced similarly to pALG2-ΔDIV. However, pALG2-ΔDIVTL, with the majority of DIV deleted except for the DIV stem and a few extra bases to form a loop, had about 10% wild type splicing, similar to those of pALG2-ΔDIVa and ΔDIVa/b. I suspected that a stem-loop structure in DIV was important for splicing. pALG2-ΔDIV+st was constructed by inserting a five-base pair GC stem and a TTTT tetra loop in the position of DIV in pALG2-ΔDIV (Fig 2.1c) and spliced with an efficiency similar to that of pALG2-ΔDIVTL (~ 10% wild type; not shown).

pALG2-ΔORF/ΔDIVa, pALG2-ΔORF/ΔDIVa/b, and pALG2-ΔORF/ΔDIVTL were deleted of the LtrA ORF to make pALG1-ΔORF/ΔDIVa, pALG1-ΔORF/ΔDIVa/b, and pALG1-ΔORF/ΔDIVTL, all of which did not splice *in vivo* (not shown).

Together, these findings indicate that the high-affinity binding site in DIVa is required for efficient splicing *in vivo*, but in the absence of DIVa, residual maturase-dependent splicing can still occur by direct binding of the protein to the catalytic core. DIVb also makes a significant contribution to splicing, either directly by contributing to the binding of the maturase, or indirectly by affecting the structure of the intron RNA. The DIV stem also contributes to splicing but can be replaced by a stem-loop structure with unrelated sequence.

2.4 UNIGENIC EVOLUTION ANALYSIS OF LTRA

To identify regions of LtrA required for splicing, I employed a method termed unigenic evolution in which a genetic assay is used to identify functional protein variants in a library containing random mutations (Deminoff et al., 1995). The mutations are then analyzed statistically to determine the ratio of missense to silent mutations compared to that expected based on codon degeneracy and the mutation frequency in the library. For unigenic evolution analysis of LtrA, I used error-prone PCR to create a library of mutant proteins in plasmid pELG2, a higher copy number version of pALG2, in which the LtrA protein is expressed from the downstream *cis* position. The higher copy number decreases plasmid loss during the assays. Sequencing of 36 randomly selected plasmids from the library showed that the mutation frequency was 0.78%, resulting in an average of ~10 amino acid substitutions per protein.

To identify functional protein variants, I first attempted to use FACS to sort cells harboring splicing-competent mutants by high GFP fluorescence. However, I found that the sorted cells were generally inviable or had lost fluorescence after overnight culturing, likely due to depletion of nucleotide pools and plasmid loss resulting from the relatively long (2.5 h) IPTG-induction of T7 RNA polymerase required to give sufficient signal intensity. For that reason, we developed an alternate assay in which the plated colonies

were duplicated onto nitrocellulose filters. The filters were then induced with IPTG for 4 h and scanned by using a FluorImager (Figure 2.7). Variants that exhibited high GFP fluorescence on the filters were isolated from the original plate, grown up in liquid culture, and assayed quantitatively by FACS to determine the level of splicing. This procedure enables the screening of 1000-1500 colonies per plate duplicated onto five to seven 35-mm diameter filters at ~200 colonies per filter, and it eliminates the need for culturing cells that have been induced. Variants that gave more than 50% of wild-type fluorescence in the FACS assay were classified as active and sequenced.

In total, I identified 111 LtrA variants that gave ≥ 50% wild-type fluorescence. These variants contained 622 mutations of which 369 were missense (Figure 2.8). To determine which regions of the LtrA protein are required for splicing, I calculated the mutability value, M, for a 25-amino acid-sliding window, using the formula

$$M = \frac{f_{Omis}}{f_{Emis}} - 1,$$

where f_{Omis} is the observed frequency of missense mutations and f_{Emis} is the expected frequency based on the mutation frequency of the library and codon degeneracy (Deminoff et al., 1995). Figure 2.9 shows a plot of the mutability value versus the amino acid residue at the center of the sliding window. Negative values in the plot indicate hypomutability with a maximal value of -1 indicating no missense mutations, and positive values indicate hypermutability, which were normalized by using the formula

$$\frac{1}{f_{Emis}} - 1,$$

so that all values lie between -1 and +1.

The mutability plot shows seven regions (A-G) spanning the RT and X domains that were hypomutable in functional LtrA variants, separated by regions that were less constrained. Although some of the latter regions appear to be hypermutable, χ^2

calculations indicate that in each case the difference from the expected frequency of missense mutations is not statiscally signifant (p values > 0.1). By contrast, the seven hypomutable regions are highly significant with A, B, C, E, and F having p values < 0.001 and D and G having p values < 0.01. At the C-terminus of the protein, the DNA-binding domain was less constrained, and the DNA endonuclease domain showed an essentially random ratio of silent to missense mutations. Additionally, several splicing-competent mutants were found to have premature stop codons, the farthest upstream being at amino acid 522 in the DNA-binding domain. These findings indicate that the RT and X domains but not the DNA-binding or DNA endonuclease domains are required for efficient splicing.

HIV-1 RT, whose structure has been determined by X-ray crystallography, is a heterodimer, with the active p66 subunit consisting of fingers, palm, thumb, connection, and RNase H domains (Kohlstaedt et al., 1992). The palm contains the RT active site, including the highly conserved YADD motif, which binds catalytically essential metal ions, while the fingers, thumb, and connection domain contribute to template and primer binding (Huang et al., 1998). Conserved RT motifs 2 and 4 in the fingers help position the template RNA at the active site and are part of the "template grip", while RT-7 in the palm interacts with primer terminus and is referred to as the "primer grip" (Jacobo-Molina et al., 1993).

In the LtrA protein, the hypomutable regions A to E correspond to the fingers and palm of HIV-1 RT, with the most conserved regions, B, C, and E, encompassing the template and primer grip regions (RT-2, -4, and -7; amino acid residues involved in template and primer binding denoted by "*T*" and "*P*", respectively, in Figure 2.8). The additional conserved regions A and upstream part of B at the N-terminus of the RT domain, include RT-0 and -1, and may be part of an extended fingers region proposed to

be involved in specific binding of the template RNA in non-LTR retroelement RTs (Chen & Lambowitz, 1997; Bibillo & Eickbush, 2002). By contrast, the catalytically essential YADD motif, which is part of RT-5 in the palm, was mutated in some variants, as were amino acid residues involved in dNTP binding in HIV-1 RT (denoted by "*N*" in the Figure). These findings extend previous results for the yeast aI2 intron, which showed that mutations in the conserved YADD motif abolish RT activity, but leave substantial splicing activity *in vivo*, indicating that RT activity *per se* is not required for splicing (Moran et al., 1995).

Domain X, which corresponds in position to the thumb and connection domains of HIV-1 RT, contains two hypomutable regions F (amino acid residues 399-429) and G (amino acid residues 461-481), which appear to be separated by a short non-conserved region. All of the amino acid residues that are conserved in the X domains of other group II IEPs were either invariant or replaced only by amino acid residues found at the same position in other group II IEPs. The invariant amino acid residues include S462, which was mutated in a splicing-defective yeast aI2 maturase (Moran et al., 1994).

The predicted secondary structure of domain X in LtrA and other group II intron maturases has three predicted α-helices in an arrangement similar to that found in the thumb of HIV-1 RT (PHD program; http://cubic.bioc.columbia.edu/PP; Jacobo-Molina et al., 1993; Blocker et al., 2005). The first two α-helices lie in hypomutable region F and the third lies in hypomutable region G. We also note that domain X is separated from the RT domain by a non-conserved region of 25 amino acid residues (positions 363-388) suggestive of an inter-domain linker.

Considered together, the results of the unigenic evolution analysis show that regions of both the RT and X domains are required for RNA splicing activity, and the critical regions include those involved in template and primer binding in other RTs. In

general, missense mutations in these regions could inhibit RNA splicing directly, or indirectly by affecting the structure of another region or by leading to a decreased amount of active protein.

2.5 BIOCHEMICAL ANALYSIS OF RT AND X DOMAIN MUTANTS

Three other people contributed to this part of the work. Dr. Manabu Matsuura generated figures 2.10 to 2.12. Dr. Hongwen Ma and Ms. Qin Wang constructed a number of mutants and conducted some initial biochemical experiments.

To investigate how the RT and X domains contribute to splicing, we constructed a series of LtrA mutants for biochemical analysis. Four of the mutants have different-sized N-terminal truncations, while the remainder have small numbers of alanine substitutions in critical regions identified in the unigenic evolution analysis. The mutant proteins were expressed in *E. coli*, purified, and assayed for RT activity and for their ability to splice and bind the Ll.LtrB intron (Figures 2.10-12). The RT assays were carried out both with the artificial template-primer substrate poly(rA)/oligo$(dT)_{18}$ and with a previously described natural substrate mimic denoted Ll.LtrB/E2+10 (Wank et al., 1999; San Filippo & Lambowitz, 2002). The latter consists of Ll.LtrB RNA (*in vitro* transcript containing the Ll.LtrB-ΔORF intron and flanking exons) with an annealed 20-mer DNA primer (E2+10), whose 3' end corresponds to that of the cleaved bottom strand normally used as the primer for TPRT. The RNA splicing assays were carried out by incubating LtrA protein with Ll.LtrB RNA at 1:1 and 10:1 protein:RNA ratios to help distinguish effects due to decreased binding affinity of the LtrA protein. The binding of the LtrA protein to Ll.LtrB RNA was assayed by equilibrium binding, and supplemented by k_{off} determinations for some mutants.

The results for the wild-type LtrA protein are in close agreement with those obtained previously (Saldanha et al., 1999; Wank et al., 1999; Matsuura et al., 2001). The

wild-type protein has high RT activity with both the poly(rA)/oligo(dT)$_{18}$ and Ll.LtrB/E2+10 substrates (Figure 2.10), promotes splicing of the Ll.LtrB intron in reaction medium containing 5 mM Mg^{2+}, where the intron is not self-splicing (Figure 2.11), and binds the Ll.LtrB RNA with high affinity (Figure 2.12). As reported previously, the apparent K_d ($K_{1/2}$) measured by equilibrium binding (2.0 nM) is substantially higher than that measured kinetically as k_{off}/k_{on} (0.12 pM) (Saldanha et al., 1999). This difference could reflect either a multi-step binding mechanism in which a stable intermediate forms prior to the final complex (Webb et al., 2001) or that the apparent K_d measured by equilibrium binding is limited by some other factor, such as the dissociation of LtrA dimers, instability of the protein at low concentrations, or long incubation times required for equilibrium at low concentrations of the interacting molecules (Saldanha et al., 1999). In either case, the k_{off} value provides the best measure of the stability of the final active complex.

First, analysis of N-terminal truncations showed that mutant proteins lacking the first 10 or 20 amino acid residues (ΔN10 and ΔN20) retained some activities, while mutant proteins with larger truncations (ΔN30 or ΔN40) were poorly expressed and lacked RT and RNA splicing activities (not shown). Strikingly, both the ΔN10 and ΔN20 mutants had substantial RT activity with the artificial template-primer substrate poly(rA)/oligo(dT)$_{18}$ (50-55% wild type), but completely lacked RT activity with the Ll.LtrB/E2+10 substrate (Figure 2.10). Further, both mutants had strongly decreased binding affinity for Ll.LtrB RNA ($K_{1/2}$'s > 200 nM; Figure 2.12(a)), but retained surprisingly high levels of RNA splicing activity (Figure 2.11(a)). This biochemical phenotype is expected for mutant proteins that have impaired interaction with the high-affinity binding site DIVa. Such mutants are expected to have strongly decreased binding affinity for the intron RNA, but may still promote splicing by binding directly to the

catalytic core (Wank et al., 1999; Matsuura et al., 2001). Further, since binding to DIVa positions the protein to initiate reverse transcription in the 3' exon of the Ll.LtrB/E2+10 substrate (Wank et al., 1999) impaired interaction with DIVa could also account for the differential inhibition of reverse transcription of that substrate relative to poly(rA)/oligo(dT)$_{18}$. The data for these mutants demonstrate that very tight binding of LtrA to the intron RNA is not essential for RNA splicing, at least under *in vitro* conditions where the protein concentration is relatively high.

The mutant DK has two amino acid substitutions (D107A, K108A) in the template grip (RT-2). This mutant had strongly decreased RT activity with both the poly(rA)/oligo(dT)$_{18}$ and Ll.LtrB/E2+10 substrates (Figure 2.10). However, the mutant retained high RNA splicing activity (Figure 11(b); the somewhat decreased amplitude at 1:1 protein:RNA likely reflects a decreased proportion of active protein) and high affinity for Ll.LtrB RNA ($K_{1/2}$ = 3.2 nM compared to 2.0 nM for wild type; Figure 2.12(b); and k_{off} = 3.1 x 10^{-4} min^{-1} compared to 3.6 x 10^{-4} min^{-1} for wild type; Figure 2.12(d)). Given the location of the mutations, a reasonable possibility is that they do not strongly affect the specific binding of Ll.LtrB RNA but rather the positioning of the RNA template at the active site. Other template and primer grip mutants tested (S218A, P219A in RT-4 and R355A, F356A, L357A, and G358A in RT-7) were poorly expressed, preventing biochemical analysis.

The mutants YADD→YAAA, at the RT active site, and R98A and L100A, which correspond to amino acid residues involved in dNTP binding in HIV-1 RT, had strongly decreased RT activity with both the poly(rA)/oligo(dT)$_{18}$ and Ll.LtrB/E2+10 substrates, as expected (Figure 2.10). However, all three mutants retained high RNA splicing activity (Figure 2.11(b)) and bound Ll.LtrB as well as the wild-type protein ($K_{1/2}$'s = 0.6 to 2.3 nM; Figure 2.12(b); k_{off} values = 2.2 to 3.1 x 10^{-4} min^{-1}; Figure 2.12(d)). These

findings show that neither RT activity nor the integrity of the RT active site is required for RNA splicing activity.

Finally, the domain X mutant EYSC (E460A, Y461A, S462A, C463A) had low RT activity with poly(rA)/oligo(dT)$_{18}$ (15% wild type) and no detectable RT activity with Ll.LtrB/E2+10 (Figure 2.10). The mutant also had no detectable RNA splicing activity at either 1:1 or 10:1 protein:RNA ratio (Figure 2.11). However, in equilibrium binding assays, the EYSC protein bound Ll.LtrB as well as the wild-type protein ($K_{1/2}$ = 0.91 nM; Figure 2.12(c)). Further, this binding was verified to be specific, as judged by failure of either the wild-type or EYSC protein to bind a group I intron RNA (Figure 2.12(c)). The k_{off} for the EYSC protein was about 30-fold faster than that for the wild-type protein (1.1 x 10^{-2} min^{-1} compared to 3.6 x 10^{-4} min^{-1}; Figure 2.12(d)), but comparable to that for LtrA bound to a small RNA containing the DIVa high affinity binding site (1.2 to 1.3 x 10^{-2} min^{-1} under the same temperature and ionic conditions; Wank et al., 1999). Together, these findings show that the EYSC mutant still binds Ll.LtrB specifically and with relatively high affinity, possibly via interactions with DIVa, but is likely defective in one or more critical weaker interactions with catalytic core regions required for splicing. Other domain X mutants tested (Y442A, Y443A, K470A, and K472A) were too poorly expressed for biochemical analysis.

2.6 DISCUSSION

Here, I developed an *E. coli* genetic assay for maturase-promoted group II intron splicing by linking the splicing of the *L. lactis* Ll.LtrB intron to the expression of GFP. To accommodate the requirement for specific 5'- and 3'-exon sequences that base pair with the intron RNA, the intron and short flanking exon sequences from the *ltrB* gene were inserted at the N-terminus of the reporter, so that splicing of the intron leads to the expression of GFP with a small N-terminal extension. The same approach should be

generally applicable to other selectable or screenable markers that tolerate N-terminal fusions. The *E. coli* genetic assay enabled us to investigate several key aspects of maturase-promoted group II intron splicing *in vivo*.

First, I found that the maturase functions most efficiently when expressed in *cis*, either from its normal location within the intron or from a position just downstream of the 3'-exon/GFP marker. The maturase also functions when expressed in *trans* from a separate plasmid, but the extent of splicing measured by GFP fluorescence was < 10% of that for either *cis* configuration. The decreased splicing efficiency when LtrA is expressed in *trans* appears to reflect both the instability of the free protein and more rapid turnover of unspliced precursor RNA when the formation of stable complex is delayed. In Ll.LtrB and other bacterial introns, the initial binding of the maturase may occur prior to the completion of transcription or translation, facilitated by the proximity of the newly synthesized N-terminus to its putative binding site in DIVa, which contains the initiation codon. The initial binding to DIVa could then nucleate other interactions with the catalytic core. Since LtrA binds the intron RNA as a dimer (Saldanha et al., 1999), the presumption is that stable binding requires at least two rounds of translation. Consistent with our results, Zhou et al. (2000), using an *L. lactis* assay in which splicing of the intron from the *ltrB* relaxase gene in one plasmid is required for conjugation of a second plasmid carrying a genetic marker, observed a 27-fold decrease in conjugation efficiency when the LtrA ORF was deleted from its normal location in DIV and expressed in *trans* from a third plasmid. However, they could not exclude that the decrease was due to the introduction of the third plasmid or variability in the assays.

Experiments analyzing the effect of deleting different intron domains or the intron ORF showed that *in vivo* splicing of Ll.LtrB intron is dependent on both the integrity of the group II intron core, whose ribozyme activity catalyzes splicing, and on expression of

active maturase, which promotes RNA folding. The intron was not detectably self-splicing *in vivo*. In agreement with previous *in vitro* assays (Wank et al., 1999), I found that the high-affinity maturase binding site in DIVa is required for efficient splicing of the Ll.LtrB intron *in vivo*, but in the absence of DIVa, 6-10% residual splicing can occur by direct binding of the maturase to the catalytic core. Thus, both the DIVa-dependent and DIVa-independent splicing mechanisms defined by previous *in vitro* assays with the Ll.LtrB intron are physiologically relevant. In an intron containing DIVa, the deletion of the peripheral regions DIVb1/b2 inhibited *in vivo* splicing by ~30%, while the complete deletion of DIVb inhibited splicing by ~60%, in reasonable agreement with *in vitro* results (Wank et al., 1999). Recent studies with the yeast aI2 intron showed that deletion of DIVb strongly inhibits maturase-promoted reverse splicing in an *in vitro* assay, but its effect was not assessed *in vivo* (Huang et al., 2003). In both the yeast intron and Ll.LtrB, DIVb could contribute to maturase-promoted reactions either by providing additional binding sites for the LtrA protein, or indirectly by affecting the conformation of DIVa or other regions of the intron RNA.

For the yeast aI2 intron, DIVa is also required for stable binding of the maturase, but its deletion inhibited *in vivo* splicing by only ~30%, a smaller effect than found here for the Ll.LtrB intron (Hunag et al., 2003). This smaller effect could reflect different relative affinities of the maturase for different regions of the intron RNA, a smaller effect of DIVa deletion on the underlying ribozyme activity, or different assays used to detect *in vivo* splicing (*e.g.*, a lower rate of turnover of spliced product would result in less apparent inhibition of RNA splicing). Notably, for both the lactococcal and yeast introns, the deletion of DIVa has a much larger effect on intron mobility than it does on RNA splicing (mobility decreased by $>10^5$ for Ll.LtrB and to undetectable levels using a less sensitive assay for the yeast aI2 intron) (Huang et al., 2003; D'Souza & Zhong). These

findings suggest that binding to DIVa is more important for intron mobility than for RNA splicing and that somewhat different RNA-protein interactions may be used for the two processes. The larger effect deletion of DIVa on intron mobility may reflect that binding to DIVa is particularly critical for positioning the protein on the Ll.LtrB RNA for initiation of reverse transcription (Wank et al., 1999).

Interestingly, the domain IV stem can be replaced by an unrelated stem-loop and splicing is restored to the level of ΔDIVa. It should be interesting to see whether adding back DIVa to the unrelated stem loop could further restore the activity.

The unigenic evolution analysis identified seven regions (A-G) spanning the RT and X domains that are hypomutable in splicing-competent LtrA variants. Hypomutable regions A-E span regions corresponding to the fingers and palm of retroviral RTs, while hypomutable regions F and G are in domain X, which likely corresponds to the thumb of retroviral RTs. By contrast, the C-terminal DNA-binding and DNA endonuclease domains were not highly constrained in splicing-competent LtrA variants and were in fact truncated by nonsense mutations in some variants. These findings together with previous *in vivo* results for the yeast aI2 intron (Zimmerly et al., 1995b) show that the DNA-binding and DNA endonuclease domains are not essential for RNA splicing.

Within the RT domain, the hypomutable regions A and the upstream part of B encompass the N-terminus of the protein and include conserved sequence motifs RT-0 and RT-1. The larger N-terminal region in non-LTR-element RTs has been postulated to be part of an extended fingers domain involved in specific binding of the template RNA for initiation of reverse transcription in non-LTR-element RTs (Chen & Lambowitz, 1997; Bibillo & Eickbush, 2002). Biochemical assays support this hypothesis by showing that the N-terminal truncation mutants ΔN10 and ΔN20 have strongly decreased binding affinity for Ll.LtrB RNA and are unable to initiate reverse transcription on Ll.LtrB RNA

using a 3'-exon primer, but retain substantial RT activity with the artificial substrate poly(rA)/oligo(dT)$_{18}$ (Figure 10). Further, both mutants retain high RNA splicing activity. These findings suggest that the N-terminal truncations specifically affect interactions with DIVa, which are required for high-affinity binding and positioning the RT to initiate reverse transcription in the 3' exon of Ll.LtrB RNA, without impairing critical but weaker interactions with catalytic core regions required for RNA splicing.

Within the RT domain, the hypomutable regions A and the upstream part of B encompass the N-terminus of the protein and include conserved sequence motifs RT-0 and RT-1. The larger N-terminal region in non-LTR-element RTs has been postulated to be part of an extended fingers domain involved in specific binding of the template RNA for initiation of reverse transcription in non-LTR-element RTs (Chen & Lambowitz, 1997; Bibillo & Eickbush, 2002). Biochemical assays support this hypothesis by showing that the N-terminal truncation mutants ΔN10 and ΔN20 have strongly decreased binding affinity for Ll.LtrB RNA and are unable to initiate reverse transcription on Ll.LtrB RNA using a 3'-exon primer, but retain substantial RT activity with the artificial substrate poly(rA)/oligo(dT)$_{18}$ (Figure 10). Further, both mutants retain high RNA splicing activity. These findings suggest that the N-terminal truncations specifically affect interactions with DIVa, which are required for high-affinity binding and positioning the RT to initiate reverse transcription in the 3' exon of Ll.LtrB RNA, without impairing critical but weaker interactions with catalytic core regions required for RNA splicing.

Hypomutable regions B, C, and E in the unigenic evolution analysis encompass conserved sequence blocks RT-2 and -4, which are part of the template grip, and RT-7, which corresponds to the primer grip. Within these regions, there was strong conservation (immutability or conservative changes) of most amino acid residues corresponding to those involved in template and primer binding in HIV-1 RT (see Figures 2.8 and 2.9). In

biochemical assays, the double mutation D107A, K108A in RT-2, which is part of the template grip, strongly inhibited RT activity with both the poly(rA)/oligo(dT)$_{18}$ and Ll.LtrB RNA/E2+10 substrates, while having little effect on intron RNA splicing or binding. Given the location of the mutations, this phenotype likely results from the disruption of one or more critical contacts required for positioning template RNA at the RT active site. Such contacts could also contribute to RNA splicing, but are expected to be weak and non-specific, since the template RNA must translocate through the active site. Thus, their disruption may have little effect on intron RNA binding, so long as other contacts, including those with the high-affinity binding site DIVa, are maintained.

Although regions of the RT domain involved in template and primer binding are required for efficient RNA splicing, the integrity of the RT active site itself is not required. Thus, a number of mutations in the unigenic evolution analysis occurred in the highly conserved YADD motif at the RT active site and in amino acid residues corresponding to those involved in dNTP binding in retroviral RTs. The biochemical assays confirmed that the mutation YAAA at the RT active site, as well as the mutations R98A and L100A, which are expected to affect dNTP binding, strongly inhibited RT activity with both the poly(rA)/oligo(dT)$_{18}$ and Ll.LtrB/E2+10 substrates, while leaving high RNA splicing activity *in vitro*. These results extend previous findings for the yeast aI2 protein, which showed that the YADD mutant lacks RT activity, but has substantial RNA splicing activity *in vivo* (Moran et al., 1995).

Finally, domain X, which corresponds in position to the thumb and connection domains of retroviral RTs, contains two hypomutable regions, F and G. A secondary structure prediction program indicates that the two hypomutable regions contain three predicted α-helices, which correspond in size and position to α-helices in the thumb domain of HIV-1 RT (Jacobo-Molina et al., 1993; Blocker et al., 2005). In an X-ray

crystal structure of HIV-1 RT containing a bound DNA template:primer and dNTP, the thumb and connection domains interact with the phosphodiester backbone of the DNA template/primer duplex (Huang et al., 1998), but during initiation the same regions are thought to interact with the template RNA/$tRNA^{Lys}$ duplex, which may be more akin to the binding of a structured group II intron RNA (Isel et al., 1999). In the unigenic evolution analysis, all of the amino acid residues that are highly conserved in the X domains of other group II IEPs were either invariant or replaced only by amino acid residues found at the same position in other group II IEPs. Further, the invariant amino acid residues include the conserved S462, which is located in one of the predicted α-helices and was mutated in a splicing-defective yeast aI2 maturase (Moran et al., 1994). Additionally, biochemical assays showed that the LtrA domain X mutant EYSC→A, which includes S462, has exactly the phenotype expected for impairment of a critical secondary contact with the catalytic core. This mutant still binds Ll.LtrB RNA specifically and with relatively high affinity, likely at least in part via interaction with DIVa, but is totally unable to promote splicing.

Considered together, our results indicate that group II IEPs function in splicing by using the RT and X domains to bind different regions of the intron RNA. The biochemical analysis of LtrA mutants suggests further that regions at the N-terminus of the RT domain interact with the high-affinity binding site in DIVa, while domain X and possibly other regions of the RT domain make secondary contacts with conserved catalytic core regions. Remarkably, the biochemical data for the N-terminal truncation mutants ΔN10 and ΔN20 show that splicing activity does not require very tight binding of the maturase to the intron RNA. Conversely, the biochemical data for the domain X mutant EYSC→A show that a mutant protein can bind Ll.LtrB tightly and specifically

but be totally unable to promote splicing, presumably due to impairment of a weaker, but critical interaction with the catalytic core.

The suggestion that the N-terminal part of the RT domain interacts with DIVa, while other parts of the RT and X domains interact with conserved catalytic core regions agrees well with findings for other introns. First, studies with the yeast aI1 and aI2 introns, which are 50% identical, showed that genetically isolated hybrids consisting of the N-terminal region of the aI1 maturase (RT-0 to -2 or -5) fused to the C-terminal region of the aI2 maturase functioned as an aI1-specific splicing factor (Anziano & Butow, 1991). These results indicate that the N-terminal region of the protein, which we find is required for binding DIVa, is engaged in intron-specific interactions, while downstream regions of the aI2 protein function equally well with aI1 and must therefore interact with regions conserved between the two introns. DIVa binding has been shown directly to contribute to intron specificity for the yeast aI1 and aI2 maturases (Huang et al., 2003). Second, cp MatK proteins, which are thought to function in splicing multiple group II introns, lack RT motifs 0-4, encompassing the region suggested to be involved in intron-specific interactions with DIVa, but retain recognizable RT motifs 5-7 along with domain X, which are suggested to interact with conserved regions of the group II intron catalytic core (Mohr et al., 1993). We note that MatK proteins still retain an N-terminal region upstream of RT 5-7, and phylogenetic analysis suggests that it has also been subject to evolutionary constraint (Young & dePamphilis, 2000). If MatK proteins do in fact function in splicing multiple group II introns, these findings imply that the N-terminal region of MatK proteins also has some function, either contributing directly to RNA binding for one or more introns, or being required for proper folding of other protein regions that function directly in RNA binding.

Finally, with respect to evolution, it has been suggested that mobile group II introns evolved either by insertion of an RT from another retroelement into a pre-existing group II intron (Kennell et al., 1993) or from a non-LTR retrotransposons that acquired self-splicing activity (Curcio & Belfort, 1996). In either case, the RT must have functioned initially in intron mobility and then adapted secondarily to function in RNA splicing. Our findings support the hypothesis that the IEP adapted to function in splicing by using, at least in part, interactions that were used initially for the specific binding of the intron RNA for the initiation of reverse transcription.

2.7 SUMMARY

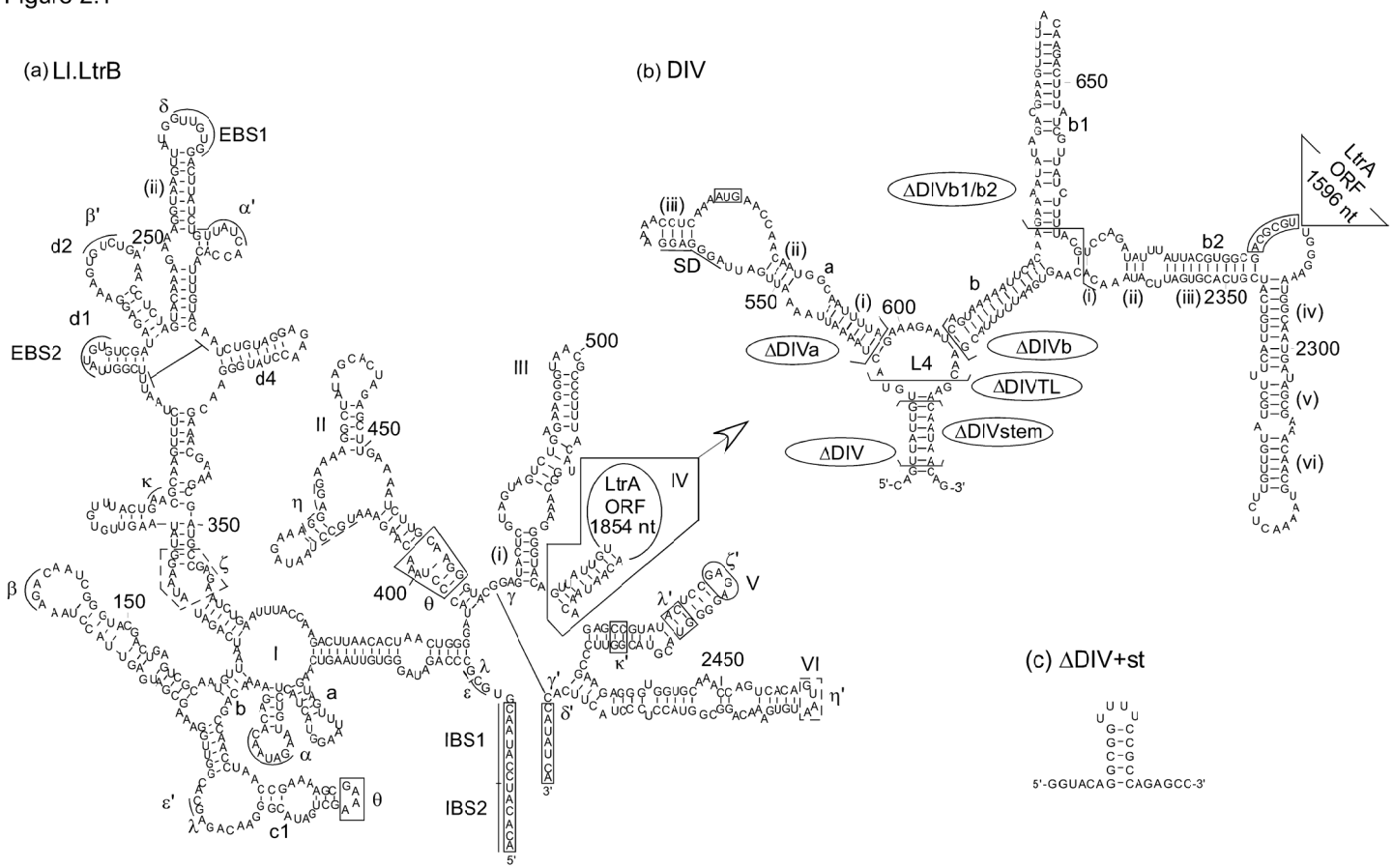
Mobile group II introns encode proteins with both reverse transcriptase activity, which functions in intron mobility, and maturase activity, which promotes RNA splicing by stabilizing the catalytically active structure of the intron RNA. Previous studies with the *L. lactis* Ll.LtrB intron suggested a model in which the intron-encoded protein binds first to a high-affinity binding site in intron subdomain DIVa, an idiosyncratic structure at the beginning of its own coding region, and then makes additional contacts with conserved catalytic core regions to stabilize the active RNA structure. Here, we developed an *E. coli* genetic assay that links the splicing of the Ll.LtrB intron to the expression of GFP and used it to study the *in vivo* splicing of wild-type and mutant introns and to delineate regions of the maturase required for splicing. Our results show that the maturase functions most efficiently when expressed in *cis* from the same transcript as the intron RNA. In agreement with previous *in vitro* assays, we find that the high-affinity binding site in DIVa is required for efficient splicing of the Ll.LtrB intron *in vivo*, but in the absence of DIVa, 6-10% residual splicing occurs by the direct binding of the maturase to the catalytic core. Critical regions of the maturase were identified by statistically analyzing ratios of missense to silent mutations in functional LtrA variants

isolated from a library generated by mutagenic PCR ("unigenic evolution"). This analysis shows that both the RT domain and domain X, which likely corresponds to the RT thumb, are required for RNA splicing, while the C-terminal D and En domains are not required. Within the RT domain, the most critical regions for maturase activity include parts of the fingers and palm that function in template and primer binding in HIV-1 RT, but the integrity of the RT active site is not required. Biochemical analysis of LtrA mutants indicates that the N-terminus of the RT domain is required for high-affinity binding of the intron RNA, possibly via direct interaction with DIVa, while parts of domain X interact with conserved regions of the catalytic core. Our results support the hypothesis that the intron-encoded protein adapted to function in splicing by using, at least in part, interactions used initially to recognize the intron RNA as a template for reverse transcription.

Figure 2.1 Predicted secondary structures of wild-type and mutant Ll.LtrB introns.

(a) The predicted secondary structure of the wild-type Ll.LtrB intron, based on the generalized group IIA intron structure in Toor *et al*. (2001). Ll.LtrB positions are numbered according to Mills *et al.* (1994). Nucleotide residues involved in the EBS/IBS pairings and tertiary interactions (Greek letters) are marked. Tertiary structure elements ζ, η, and η' are demarcated by dashed lines to indicate previously discussed uncertainties in defining these elements (Matsuura et al., 2001). 5'- and 3'-exon sequences are boxed. (b) The predicted secondary structure of DIV, containing the Shine-Dalgarno (SD) sequence and initiation codon (boxed) of the LtrA ORF. The location of the 1596 nts deleted in the ΔORF intron is indicated, and an *Mlu*I site introduced in constructing the deletion is boxed. The boundaries of intron deletion mutants (circled) are indicated by brackets. (c) The artificial stem-loop used to replace DIV.

Figure 2.1


(a) LI.LtrB
(b) DIV
(c) ΔDIV+st
EBS1
EBS2
IBS1
IBS2
LtrA ORF 1854 nt
LtrA ORF 1596 nt
ΔDIVa
ΔDIVb
ΔDIVb1/b2
ΔDIVTL
ΔDIVstem
ΔDIV
L4
SD
5'-GGUACAG-CAGAGCC-3'

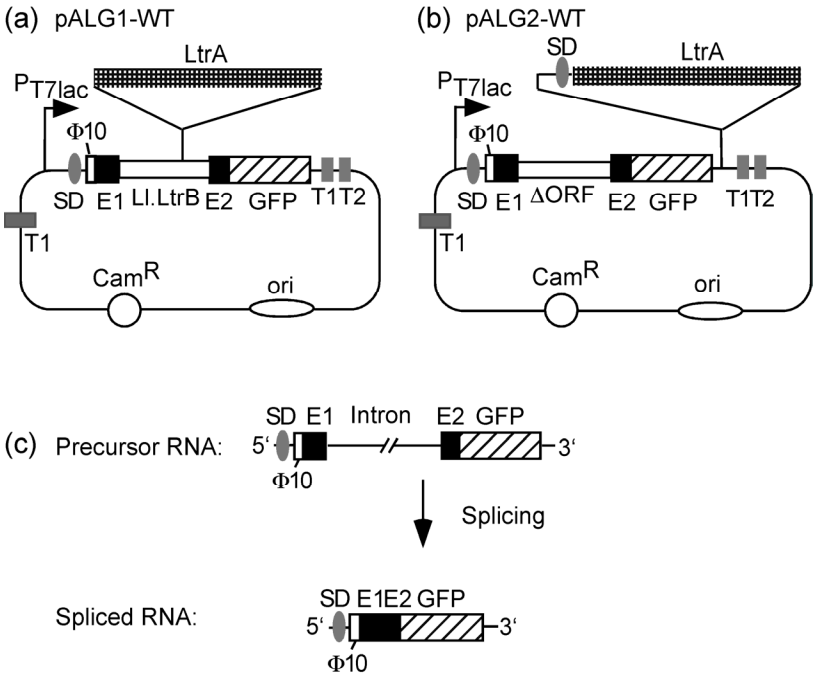


Figure 2.2 *E. coli* genetic assay of maturase-dependent splicing of the Ll.LtrB intron.

(a) and (b) Plasmids pALG1-WT and pALG2-WT, respectively. Both plasmids contain *ltrB*/GFP fusions cloned downstream of a T7lac promoter (P_{T7lac}; arrow) and ϕ10 Shine-Dalgarno (SD) sequence. The *ltrB*/GFP fusion is comprised of the first eight codons of the ϕ10 gene linked to a segment of the *L. lactis ltrB* gene consisting of the Ll.LtrB intron (open rectangle) and flanking exon sequences (E1 and E1; black boxes), with E2 linked in-frame to codon 2 of the GFP (hatched). pALG1-WT contains the full-length Ll.LtrB intron with the LtrA ORF encoded in DIV, whereas pALG2-WT contains an Ll.LtrB-ΔORF intron with the LtrA ORF cloned downstream of the GFP coding sequence. T1 and T2 are *E. coli rrnB* transcription terminators. (c) Schematic showing splicing of the Ll.LtrB intron linked to the expression of GFP. Splicing leads to the expression of GFP with 42-amino acid residue fusion corresponding to the ϕ10 sequence and ligated exons. In the absence of splicing, translation is terminated by stop codons within the intron.

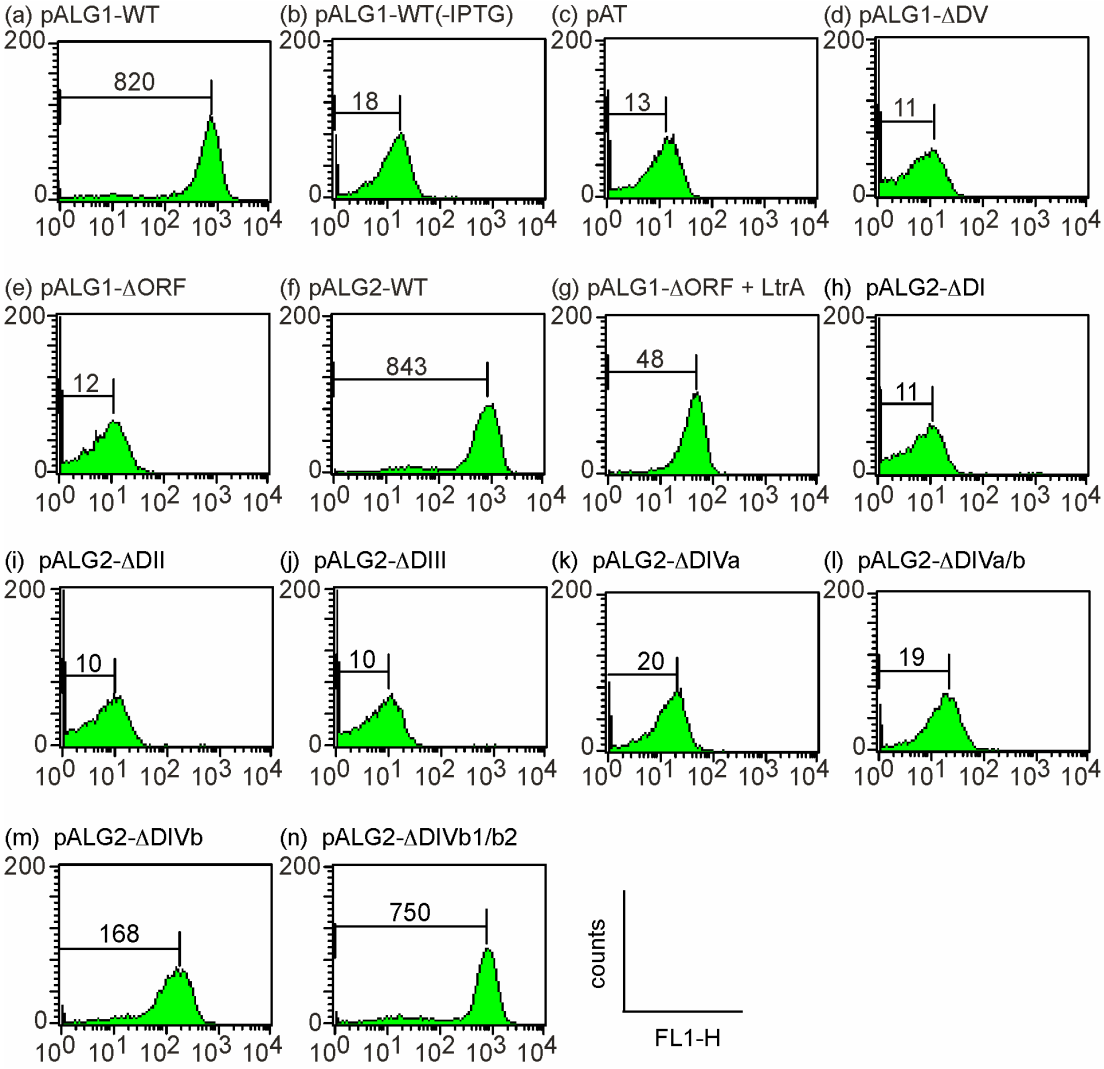


Figure 2.3 FACS assay of *in vivo* splicing of the Ll.LtrB intron.

Cells containing the indicated constructs with or without (-IPTG) induction were analyzed by flow cytometry to count the number of cells in a given window of fluorescence intensity. Peak positions are indicated by horizontal lines. In panel (g), the LtrA protein was expressed from plasmid pLI1P. Data are representative of at least two independent repeats for each construct.

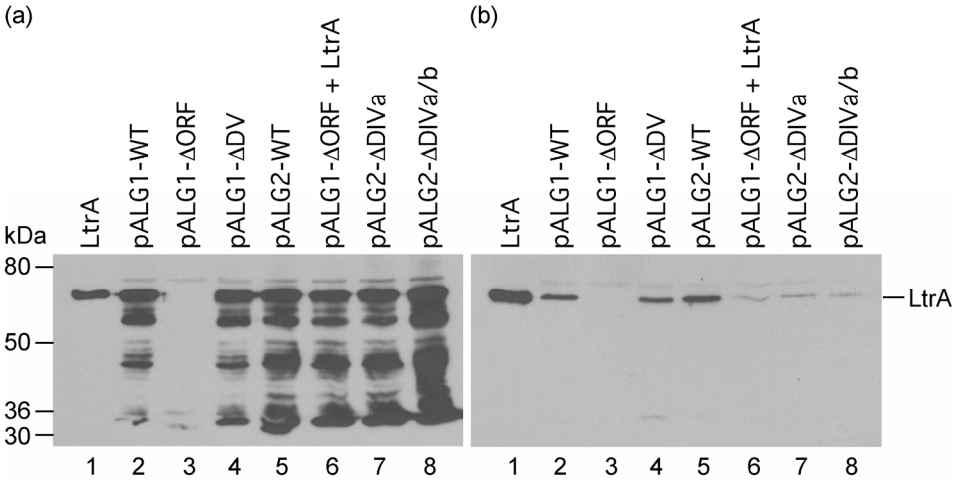


Figure 2.4 Immunoblot analysis of LtrA protein expressed from different constructs.

E. coli HMS174(DE3) containing the indicated constructs was induced with IPTG, and (a) total and (b) soluble proteins were resolved in an 8% polyacrylamide/1% SDS gel, blotted to a nitrocellulose membrane, and probed with an anti-LtrA polyclonal antibody preparation. Each gel lane was loaded with protein from 0.044 $O.D._{595}$ of cells, and equal loading was confirmed by AuroDye staining of the immunoblot (not shown). Lanes: (1) purified LtrA protein (0.03 μg); (2) pALG1-WT; (3) pALG1-ΔORF; (4) pALG1-ΔDV; (5) pALG2-WT; (6) pALG1-ΔORF plus LtrA expressed in *trans* from pET-LtrA; (7) pALG2-ΔDIVa; (8) pALG2-ΔDIVa/b. The more rapidly migrating bands in the total protein samples are degraded LtrA.

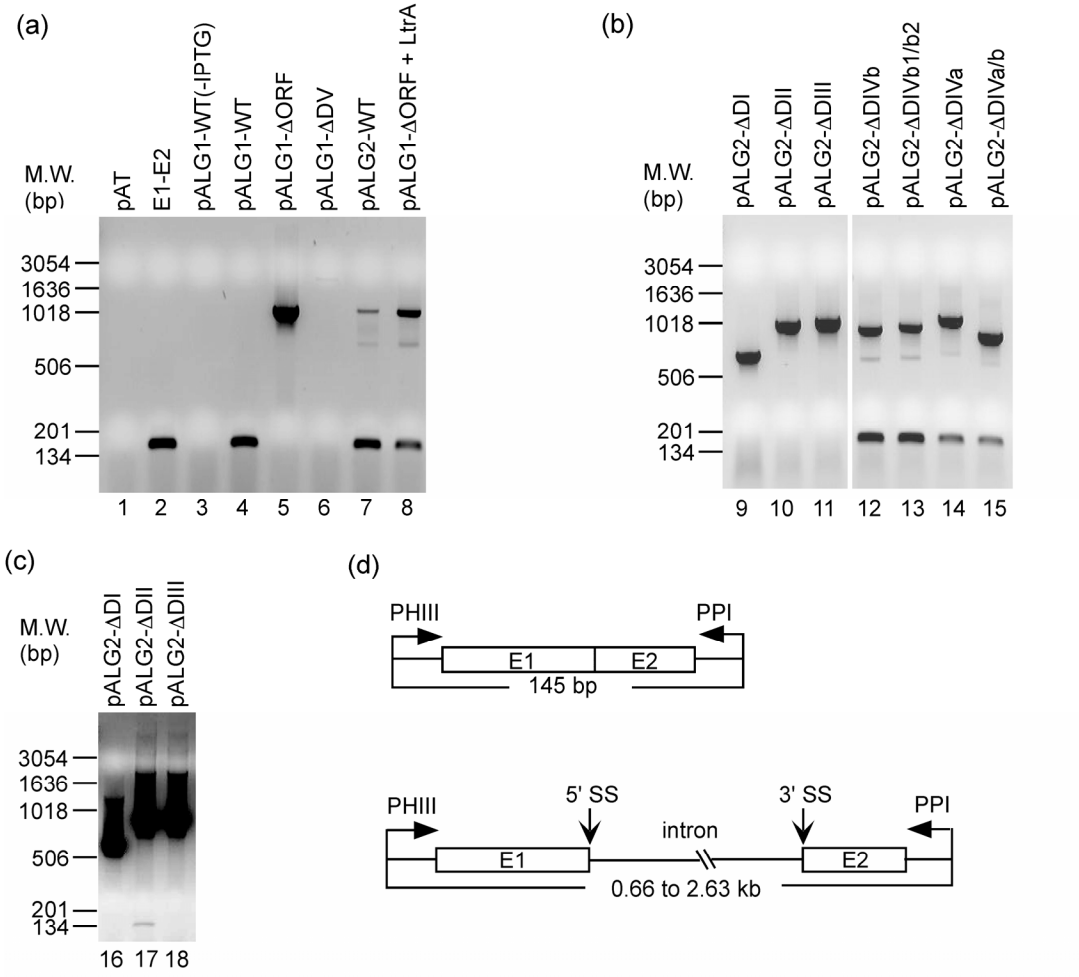
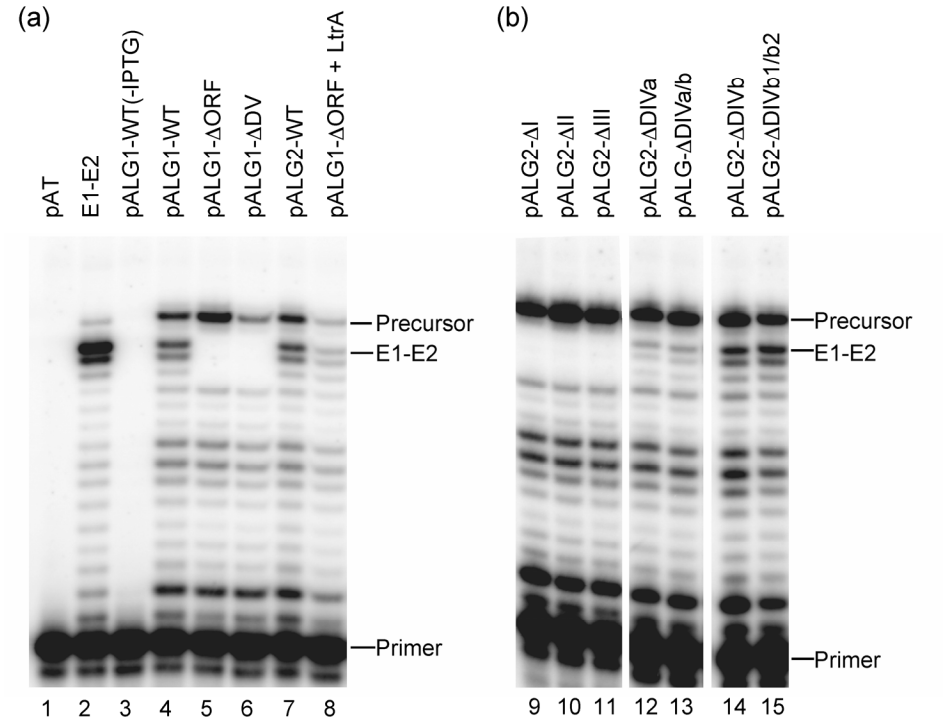


Figure 2.5 RT-PCR assay of *in vivo* splicing of the Ll.LtrB intron.

(a) to (c) RT-PCR assays. RNA from cells containing the indicated constructs was used for RT-PCR with primers PHIII and PPI, and the products were analyzed in a 1% agarose gel containing ethidium bromide (0.5 μg/ml) against 1-kb DNA size markers (Invitrogen). PCR was carried out for 16 cycles in panels (a) and (b) and 35 cycles in panel (c). (d) Schematic showing predicted RT-PCR products for unspliced precursor and ligated exon RNAs.



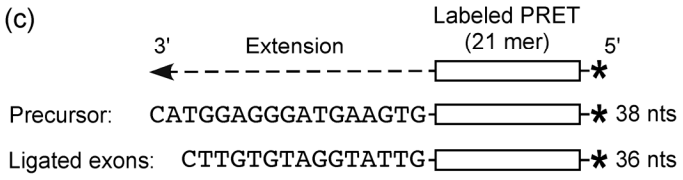


Figure 2.6 Poisoned primer extension assay of *in vivo* splicing of the Ll.LtrB intron.

(a) and (b) Poisoned-primer assays. The same RNA preparations as in Figure 5 were used for primer extension with 5'-labeled primer PRET (complementary to the first 21 nts of E2) and M-MLV RT in the presence of ddCTP to terminate reverse transcription at the first G-residue of the template. The products were analyzed in a denaturing 10% polyacrylamide gel, which was dried and quantified with a PhosphorImager. Constructs are indicated at the top of the gel. (c) Schematic showing the predicted products for unspliced precursor and ligated-exon RNAs. * indicates 5'-label.

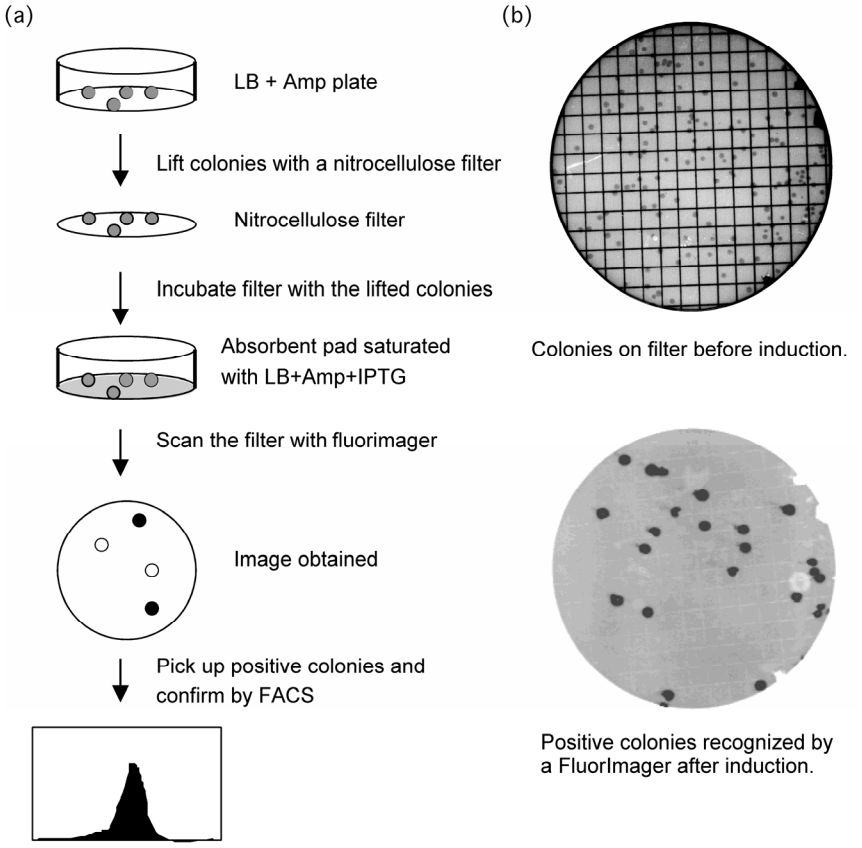


Figure 2.7 Colony-based fluorescence assay of maturase-promoted Ll.LtrB splicing linked to GFP expression.

(a) Flow chart of assay. Colonies growing on plates containing LB plus ampicillin were duplicated onto five to seven 35-mm diameter nitrocellulose filters and incubated on an absorbent pad saturated with LB + ampicillin + IPTG for 4 h at 37°C. The filters were then scanned by using a FluorImager. (b) Scanned filters. *E. coli* containing pELG2-WT (splicing competent) and pELG2-ΔORF (splicing defective due to lack of LtrA protein) were mixed at a ratio of 1:10, plated on LB containing ampicillin, and subjected to the colony-based fluorescence assay described in panel (a). The top shows an optical scan of the filter, and the bottom shows the FluorImager scan, with splicing-competent colonies that produce GFP appearing black against a light background.

Figure 2.8 Summary of missense mutations in splicing-competent LtrA variants.

Libraries of random LtrA mutants in pELG2 were screened by using the colony-based fluorescence assay to identify those that could efficiently splice the Ll.LtrB-ΔORF. Individual mutants were then retested by FACS assay, and those that had ≥ 50% wild-type splicing activity were sequenced. The Figure shows the wild-type LtrA amino acid sequence, with amino acid substitutions or premature stop codons (asterisks) found in 111 variants having ≥ 50% wild-type splicing activity indicated below. Conserved regions RT-0 to -7, and domains X, D, and En are delineated above. Consensus sequences for RT-0 to –7 (Xiong & Eickbush, 1992; Zimmerly et al., 2001) and domain X (based on current sequence alignments; F. Blocker, G. Mohr and A.M.L., unpublished data) are indicated above the LtrA sequence: b, basic; h, hydrophobic; n, nucleophilic; o, aromatic; *N*, amino acid residue involved in dNTP binding; *P*, amino acid residues involved in primer binding ("primer grip"); *T* amino acid residue involved in template binding ("template grip"). Domain X α-helices predicted by PredictProtein PHD (http://cubic.bioc.columbia.edu/pp/) were represented by cylinders.

Figure 2.8

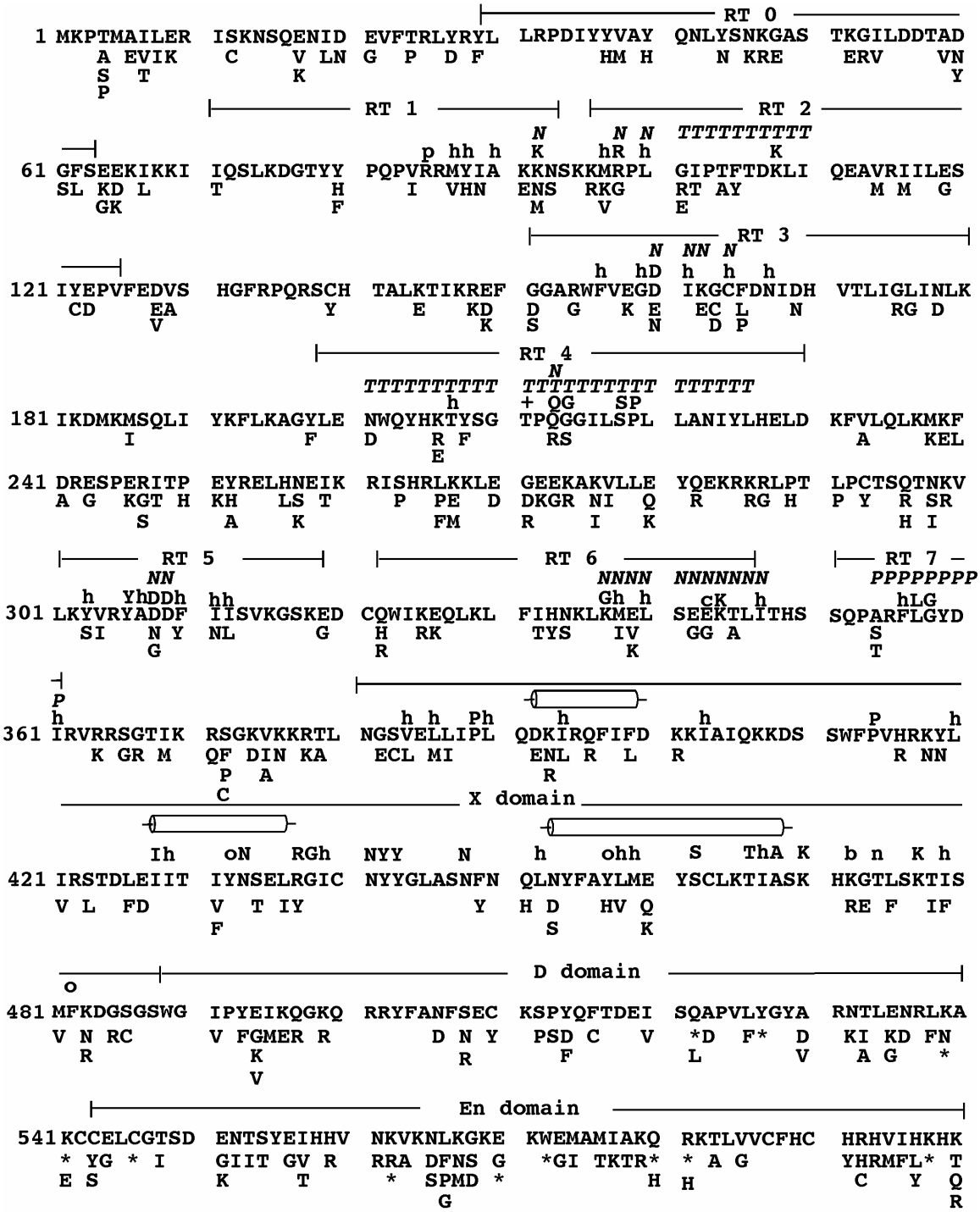
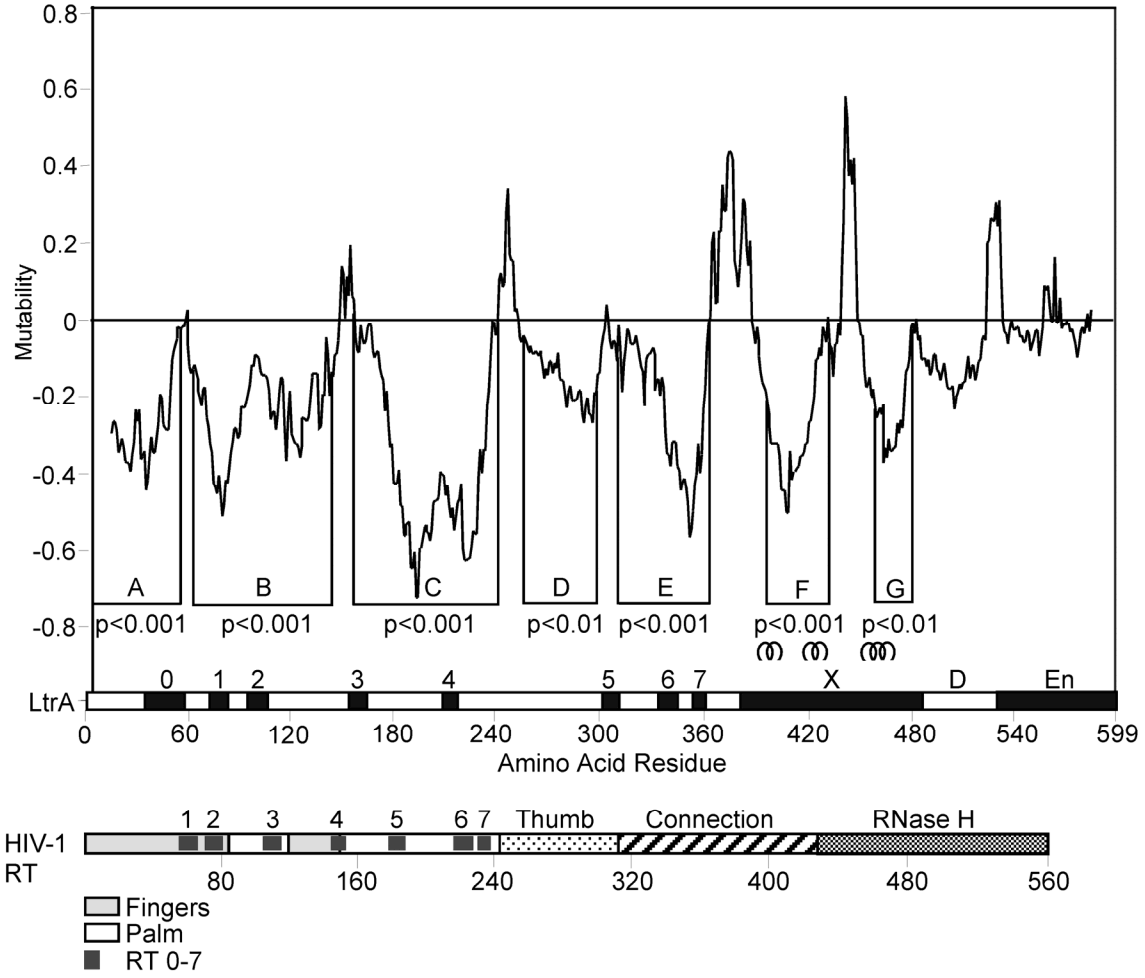


Figure 2.9 Mutability plot based on unigenic evolution analysis of LtrA variants that efficiently splice the Ll.LtrB-ΔORF intron.

The mutability value (M), which relates the frequencies of expected and observed missense mutations (f_{Omis} and f_{Emis}, respectively), was calculated for a 25-amino acid residue sliding window by using the equation $M = (f_{Omis} / f_{Emis}) - 1$ and plotted against the amino acid at the center of the window (Deminoff et al., 1995). Negative values indicate hypomutability, with a maximum value of -1 indicating no missense mutations, and positive values indicate hypermutability, which was normalized using the formula $M = (f_{Omis}/f_{Emis} - 1)/(1/f_{Omis} - 1)$, so that all values would fall between -1 and +1. χ^2 values were calculated for each hypomutable region, as described in Materials and Methods, and p values are indicated below each region. The location of conserved RT sequence motifs 0-7 and LtrA domain boundaries are shown on the X-axis, and the locations of three predicted α-helical regions in domain X are shown above. At the bottom is a schematic of HIV-1 RT with the fingers (gray), palm (open boxes), thumb (dots), connection (hatched), and RNase H (squares) domains.

Figure 2.9


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
Mutability
A
B
C
D
E
F
G
p<0.001
p<0.001
p<0.001
p<0.01
p<0.001
p<0.001
p<0.01
0
1
2
3
4
5
6
7
X
D
En
LtrA
0
60
120
180
240
300
360
420
480
540
599
Amino Acid Residue
1 2
3
4
5
6 7
Thumb
Connection
RNase H
HIV-1
RT
80
160
240
320
400
480
560
Fingers
Palm
RT 0-7

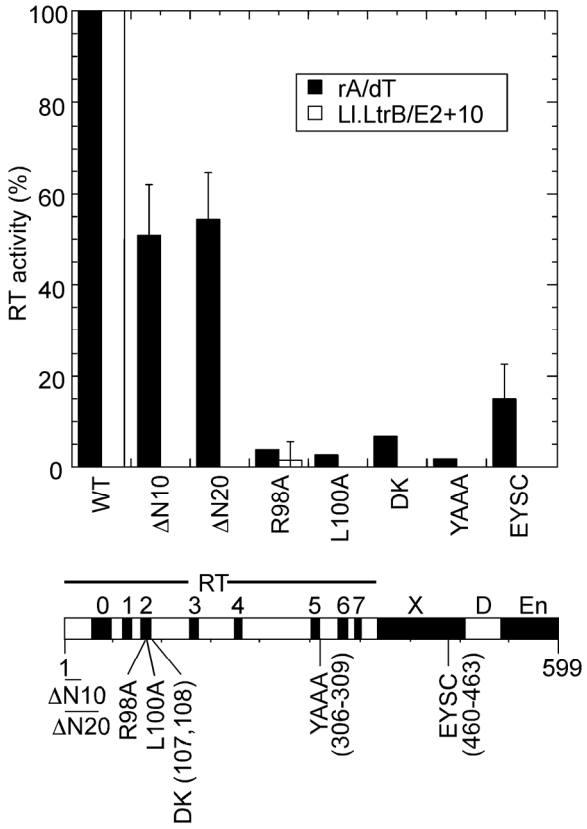


Figure 2.10 Reverse transcriptase assays.

The bar graphs show RT activity of wild-type and mutant LtrA proteins assayed with poly(rA)/oligo(dT)$_{18}$ (black bars) and Ll.LtrB/E2+10 substrates (white bars). The RT activities are expressed as percent wild-type activity assayed in parallel. The wild-type activities were ~2 x 10^6 cpm for poly(rA)/oligo(dT) and ~1.5 x 10^6 cpm for Ll.LtrB/E2+10. Data are the mean for at least three assays in each case, with standard deviations indicated by error bars. The schematic below shows the location of mutations in the LtrA protein, with protein regions delineated as in Figure 2.9.

Figure 2.11 RNA splicing assays for wild-type and mutant LtrA proteins.

For presentation, the mutants are divided into three groups: (a) N terminus RT domain; (b) RT domain; and (c) domain X. Splicing time-courses were carried out with 20 nM or 200 nM LtrA proteins and 20 nM ^{32}P-labeled Ll.LtrB RNA (1:1 and 10:1 molar ratio, respectively) at 30°C, and products were analyzed in a denaturing 4% polyacrylamide gel. Data from splicing time-courses were best fit to an equation with two exponentials to obtain k_{obs} for the fast and slow phases (k_1 and k_2, respectively). The biphasic kinetics and residual unreactive RNA under protein excess conditions are attributed to more slowly folding and unreactive RNA conformers (Saldanha et al., 1998). The numbers in parentheses indicate the proportion of input RNA that spliced in the fast and slow phases for the experiments shown. Each splicing time-course was repeated at least twice with similar results.

Figure 2.11

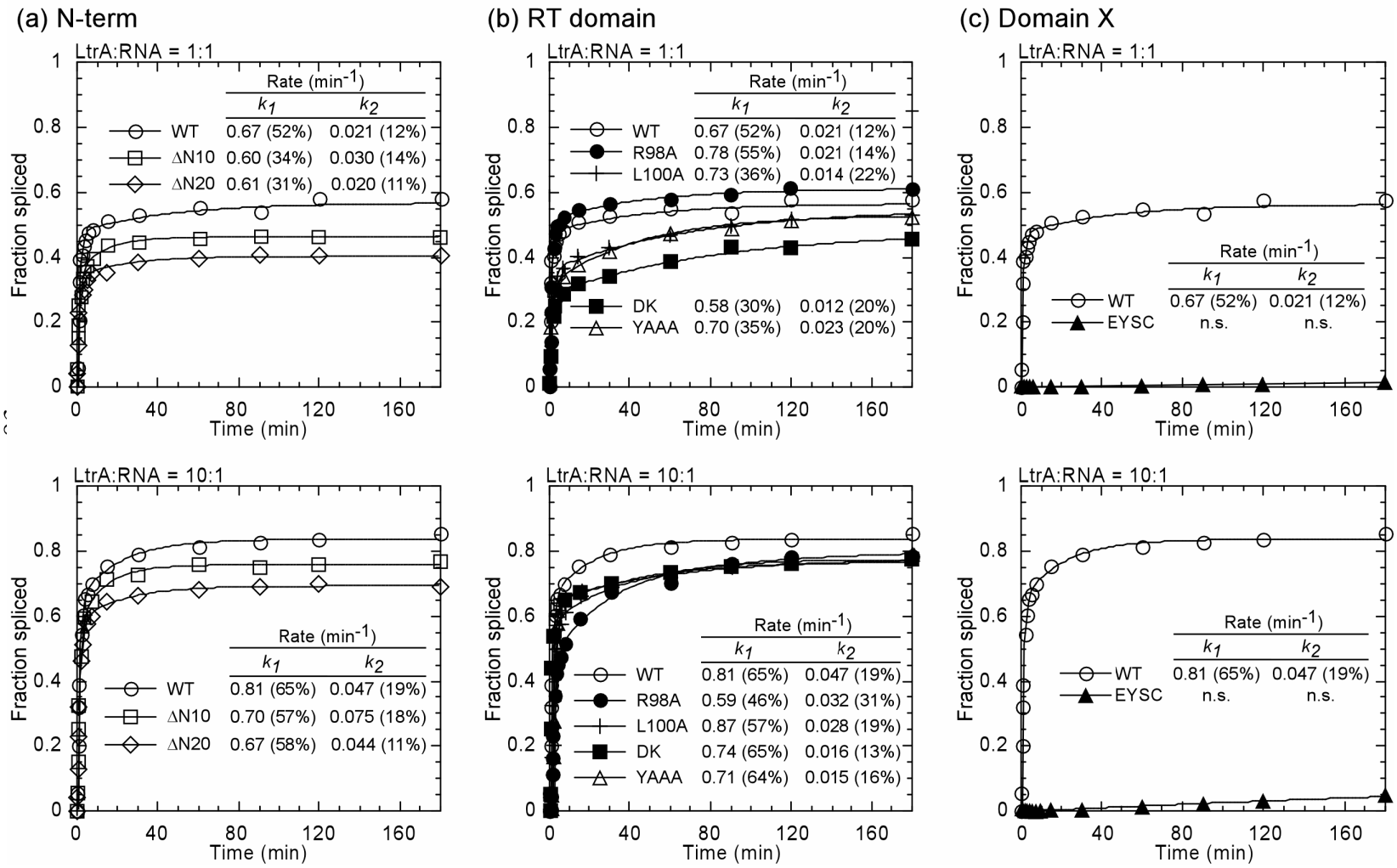
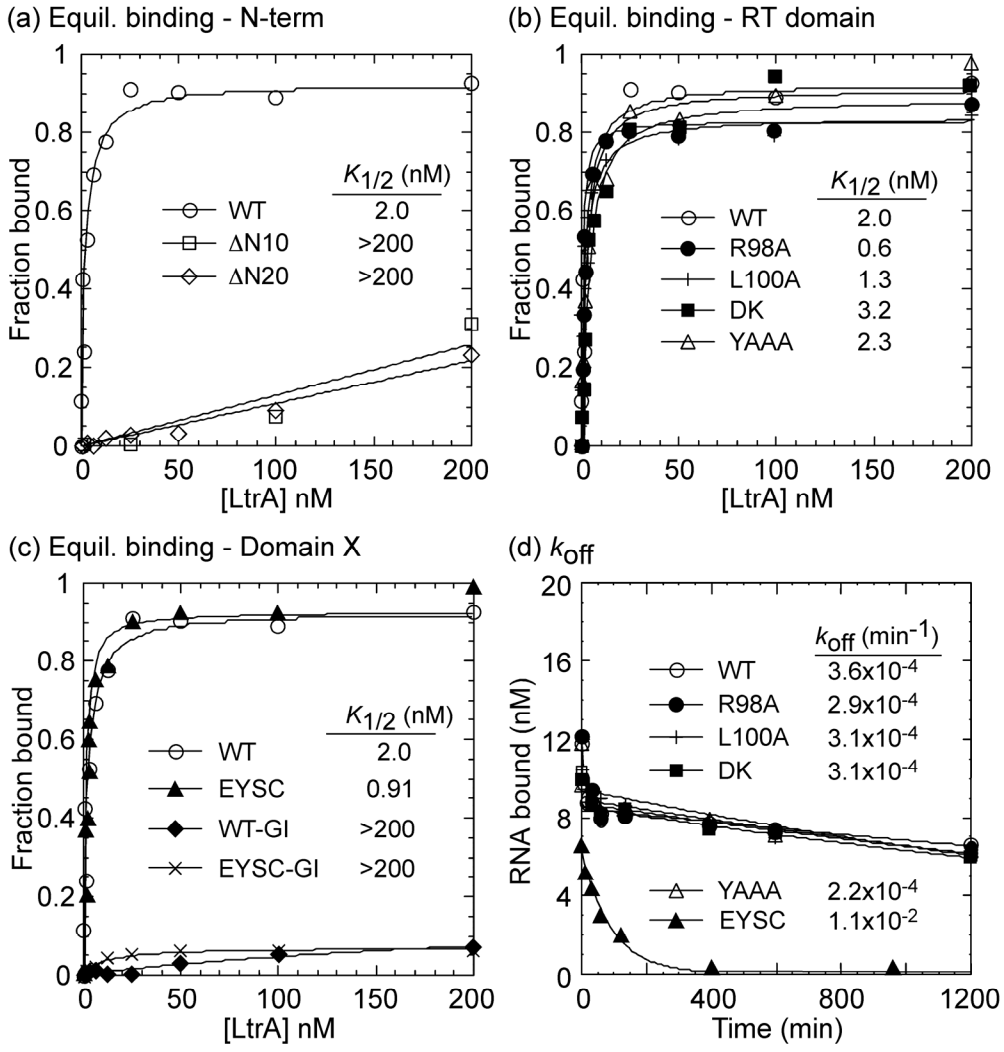

(a) N-term
(b) RT domain
(c) Domain X
LtrA:RNA = 1:1
LtrA:RNA = 10:1
Fraction spliced
Time (min)
Rate (min-1)
k1
k2
WT 0.67 (52%) 0.021 (12%)
ΔN10 0.60 (34%) 0.030 (14%)
ΔN20 0.61 (31%) 0.020 (11%)
R98A 0.78 (55%) 0.021 (14%)
L100A 0.73 (36%) 0.014 (22%)
DK 0.58 (30%) 0.012 (20%)
YAAA 0.70 (35%) 0.023 (20%)
EYSC n.s. n.s.
WT 0.81 (65%) 0.047 (19%)
ΔN10 0.70 (57%) 0.075 (18%)
ΔN20 0.67 (58%) 0.044 (11%)
R98A 0.59 (46%) 0.032 (31%)
L100A 0.87 (57%) 0.028 (19%)
DK 0.74 (65%) 0.016 (13%)
YAAA 0.71 (64%) 0.015 (16%)
EYSC n.s. n.s.

Figure 2.12 RNA binding assays.

(a)-(c) Equilibrium binding assays. Mutants are divided into three groups as in Figure 11. Equilibrium binding was carried out by incubating ^{32}P-labeled Ll.LtrB RNA (5 pM) with increasing concentrations of wild-type or mutant LtrA proteins for 1 h at 30^{o}C, and then counting radioactivity in RNP complexes retained on nitrocellulose filters. Each binding assay was repeated at least three times with similar results. In panel (c), binding of a ^{32}P-labeled group I (GI) intron RNA (the *Tetrahymena thermophila* large subunit rRNA-ΔP5abc intron; Mohr et al., 1994) was tested as a control for binding specificity. For ΔN10 and Δ20, similar weak binding was observed with incubation times ranging from 5 min to 1 h. (d) k_{off} determinations. k_{off} was measured by forming complex between 20 nM ^{32}P-labeled Ll.LtrB RNA and 20 nM LtrA protein, then diluting 20-fold into reaction medium containing an 100-fold excess of unlabeled Ll.LtrB RNA and monitoring dissociation of the complex as a function of time by nitrocellulose filter binding. The data for WT, R98A, L100A, DK, and YAAA were best fit to equations with two exponentials, with the k_{off} values given in the Figure being those for the slowly dissociating stable complex. The data for EYSC best fit an equation with a single exponential. The smaller amount of initial complex for the EYSC protein likely reflects a decreased amount of active protein.

Figure 2.12


(a) Equil. binding - N-term
Fraction bound
[LtrA] nM
$K_{1/2}$ (nM)
WT 2.0
ΔN10 >200
ΔN20 >200
(b) Equil. binding - RT domain
Fraction bound
[LtrA] nM
$K_{1/2}$ (nM)
WT 2.0
R98A 0.6
L100A 1.3
DK 3.2
YAAA 2.3
(c) Equil. binding - Domain X
Fraction bound
[LtrA] nM
$K_{1/2}$ (nM)
WT 2.0
EYSC 0.91
WT-GI >200
EYSC-GI >200
(d) k_{off}
RNA bound (nM)
Time (min)
k_{off} (min^{-1})
WT $3.6x10^{-4}$
R98A $2.9x10^{-4}$
L100A $3.1x10^{-4}$
DK $3.1x10^{-4}$
YAAA $2.2x10^{-4}$
EYSC $1.1x10^{-2}$

Chapter 3 Identification of potential binding sites for DIVa and conserved catalytic core regions in a group II intron-encoded reverse transcriptase

Mobile group II introns encode proteins that have reverse transcriptase (RT) and RNA splicing ("maturse") activities (Lambowitz et al., 1999; Belfort et al., 2002; Lambowitz & Zimmerly, 2004; Pyle & Lambowitz, 2005). The intron-encoded protein (IEP) promotes splicing by binding specifically to the intron RNA, helps fold and stabilizes the catalytically active RNA structure (Matsuura et al., 2001). It then remains bound to the excised intron RNA in RNPs that promote mobility. Mobility occurs by a remarkable mechanism in which the excised intron RNA reverse splices directly into one strand of a DNA target site, and the IEP synthesizes a cDNA copy of the inserted intron RNA (Lambowitz & Zimmerly, 2004).

The best-characterized mobile group II introns are the yeast mtDNA aI1 and aI2 introns and the *L. lactis* Ll.LtrB intron. The proteins encoded by these introns have four conserved domains, denoted RT, X, DNA-binding (D), and DNA-endonuclease (En). The RT domain contains conserved sequence motifs RT-1 to RT-7 found in the fingers and palm domains of HIV-1 RT, but like other non-LTR-retroelement RTs has a series of insertions relative to HIV-1 RT (Xiong & Eickbush, 1990; Zimmerly et al., 2001). These include an extended N-terminal region, containing conserved motif RT-0, and internal insertions, denoted RT-2a, -3a, -4a, -7a, and ti. Domain X is a site of mutations that affect RNA splicing activity and corresponds at least in part to the RT thumb. Both the RT and X (thumb) domains are involved in interactions with the intron RNA (Cui et al., 2004). Domains D and En interact with and cleave the DNA target site during intron mobility (San Filippo & Lambowitz, 2002).

Studies with the Ll.LtrB intron showed that its IEP, denoted the LtrA protein, has a high-affinity binding site in intron subdomain DIVa, an idiosyncratic stem-loop structure that contains the ribosome-binding site (RBS) and initiation codon of the LtrA ORF (Wank et al., 1999; Singh et al., 2002; Watanabe et al., 2004). In an *in vivo* assay using the *E. coli lacZ* gene as a reporter, LtrA was shown to down-regulate its own translation by binding to the RBS. The binding of LtrA to the RBS presumably also prevents further ribosome entry into the intron, which would interfere with splicing (Singh et al., 2002). DIVa contributes to the specificity of the intron/maturase interaction, but by itself is not sufficient to dictate intron specificity. Replacing DIVa of the yeast aI1 and aI2 introns with that of the Ll.LtrB intron did not enable LtrA to splice the modified aI1 and aI2 introns (Watanabe & Lambowitz, unpublished data).

In addition to binding DIVa, LtrA makes weaker but critical contacts with conserved catalytic core regions to stabilize the catalytically active RNA structure. RNA footprinting experiments identified potential secondary binding sites in DI, DII, and DVI (Matsuura et al., 2001), and LtrA forms stable complexes with DI as well as DIV (Wank et al., 1999; Rambo & Doudna, 2004). Similar findings were made for the yeast aI2 intron, whose IEP likewise has a high-affinity binding site in DIVa, but supports substantial maturase-promoted splicing in the absence of DIVa (Huang et al., 2003). In both the yeast and lactococcal introns, deletion of DIVa almost completely abolishes intron mobility (D'Souza & Zhong, 2002; Huang et al., 2003). This greater effect on intron mobility may reflect in part that the binding of the IEP to DIVa is critical for positioning the RT to initiate reverse transcription of the intron RNA (Wank et al., 1999).

I previously developed an *E. coli* genetic assay for maturase activity by linking the splicing of the Ll.LtrB-ΔORF intron (referred to in this chapter as the wild-type intron) to the expression of GFP, and I used this assay in conjunction with unigenic

evolution analysis to delineate regions of the IEP required for RNA splicing (Cui et al., 2004). Mapping of regions that are highly conserved in splicing-competent LtrA variants on a three-dimensional model of LtrA, constructed by threading on HIV-1 RT crystal structures, suggested an RNA-binding surface that includes the extended N-terminal fingers domain, regions around the template-primer binding track, and the back of the hand (Blocker et al., 2005). Biochemical analysis of mutants showed that small deletions at the N-terminus of LtrA abolished high-affinity intron binding and intron-specific RT activity, but maintained substantial splicing activity, while a domain X mutant retained high-affinity binding but completely lacked splicing activity (Cui et al., 2004). These findings suggested a model in which N-terminal regions of the RT domain interact with DIVa, while domain X interacts with conserved catalytic core regions. Consistent with this model, plant MatK proteins, degenerate maturases suggested to function in splicing multiple chloroplast group II introns (Vogel et al., 1999), have divergent RT domains, which lack recognizable N-terminal motifs RT-1 to RT-3, but retain a well-conserved domain X (Mohr et al., 1993).

In this chapter, I developed a new *E. coli* genetic assay to screen for LtrA mutants capable of promoting residual splicing of an Ll.LtrB-ΔORF intron with DIVa deleted (referred to as the ΔDIVa intron). These mutants must retain interactions with the catalytic core, but need not maintain interactions with DIVa. By using this genetic assay in conjunction with unigenic evolution analysis and three-dimensional modeling, I identified regions of LtrA that are potentially involved in binding DIVa and the catalytic core.

3.1 IDENTIFICATION OF LTRA REGIONS REQUIRED FOR RESIDUAL SPLICING OF AN LL.LTRB-ΔDIVA INTRON

To identify regions of LtrA that promote residual splicing in the absence of DIVa, I used error-prone PCR to construct a library of LtrA variants containing random mutations. The LtrA library was swapped for the LtrA ORF in the plasmid pELG2-ΔORF/ΔDIVa, which contains the ΔDIVa intron and flanking exons, with the 3' exon linked in-frame to the 2nd codon of the GFP ORF (Fig. 3.1). The LtrA protein library was expressed from a position just downstream of the 3' exon so that mutations in protein coding sequence did not affect the intron RNA structure. Splicing of the intron leads to the expression of GFP, which is otherwise aborted by multiple stop codons present in the intron sequence.

The residual splicing of the ΔDIVa intron is promoted by direct binding of LtrA to the catalytic core, so that amino acid residues required for binding DIVa need not be conserved in splicing-competent variants. I constructed a library of LtrA mutants in the ΔDIVa background. Because residual splicing in the absence of DIVa results in only low GFP fluorescence, I identified splicing-competent LtrA variants by a colony immunoblot assay, in which I used an anti-GFP antibody to monitor GFP expression in individual colonies (Fig. 3.2). In each case, the ability of a mutant protein to splice the ΔDIVa intron was confirmed by RT-PCR. I sequenced 170 splicing-competent LtrA variants that satisfied these criteria, and these contained 575 missense mutations and 370 silent mutations (summarized in Fig. 3.3).

To identify regions of LtrA required for splicing the ΔDIVa intron, I calculated the mutability value, M, for a 25-amino acid residue sliding window, using the formula:

$$M = \frac{f_{Omis}}{f_{Emis}} - 1,$$

in which f_{Omis} is the observed frequency of missense mutations, and f_{Emis} is the expected frequency based on the mutation frequency of the library and codon degeneracy. Figure 3.4 shows the resulting mutability plot for LtrA-promoted splicing of the ΔDIVa intron, with a previous mutability plot for variants that efficiently spliced the DIVa-containing wild-type intron shown in dashed lines for comparison.

Two regions (B and E) at the N-terminus of the RT domain immediately stand out as being required for efficient splicing of the wild-type intron, but not for the residual splicing of the ΔDIVa intron. Region B is centered on RT-1, while region E encompasses insertion RT-3a. Although both regions appear to be somewhat hypermutable ($M > 0$) in the ΔDIVa plot, χ^2 calculations indicate that the apparent increase in mutation frequency is not statistically significant (p values > 0.1).

Statistical tests identified two additional regions, A and H, that are also significantly constrained in variants that splice the wild-type intron, but not in those that splice the ΔDIVa intron (region A, $p<0.001$ and >0.05 for the wild-type and ΔDIVa introns, respectively; region H, $p<0.01$ and $p>0.2$ for the wild-type and ΔDIVa introns, respectively). Region A corresponds to the extended N-terminus of the protein, while region H is located in domain X and includes a small thumb-domain insertion, denoted ti (Blocker et al., 2005). Two other subregions at the beginning and end of region F appear more conserved for splicing the wild-type intron than the ΔDIVa intron, but χ^2 tests indicated that the difference is not significant (p values <0.01 for both the wild-type and ΔDIVa introns).

The mutability plots raise the possibility that some of the insertions in LtrA relative to HIV-1 RT might be involved in DIVa binding. Figure 3.5 compares mutations tolerated in each of these insertions for splicing the wild-type and ΔDIVa introns. Strikingly, insertions RT-3a and ti, contained within regions E and H, respectively, were

highly conserved in variants that splice the wild-type intron ($p<0.01$ for RT-3a and $p<0.05$ for ti), but mutable in variants that splice the ΔDIVa intron ($p>0.5$ for RT-3a and $p>0.2$ for ti). By contrast, the other insertions were either similarly constrained for splicing the wild-type and ΔDIVa introns (*e.g.*, RT-0, $p>0.2$ and RT-4a, $p>0.1$ for both introns), or less constrained for splicing the wild-type than the ΔDIVa intron (RT-2a, $p>0.1$ and $p<0.01$, and RT-7a, $p>0.2$ and $p<0.05$ for the wild-type and ΔDIVa introns, respectively).

Regions C, F, and J in the mutability plots are significantly constrained in LtrA proteins that function in splicing both the wild-type and ΔDIVa introns (p values < 0.01). These regions are either required for maintaining LtrA structure, which is essential for LtrA function, or for binding the catalytic core to promote the splicing of both introns.

Finally, several regions (D, G, I, and K) are more highly constrained in variants that splice the ΔDIVa intron than in those that splice the wild-type intron. The greater constraints on these four regions for splicing the ΔDIVa intron may reflect that the effect of disrupting weaker interactions with the catalytic core is amplified when the high-affinity binding site in DIVa is deleted. It is also possible that in the absence of the high-affinity binding site in the intron, other regions of the protein are exposed to bind the catalytic core, or structural changes in the intron RNA upon deletion of DIVa affect protein binding.

3.2 MAPPING OF CONSERVED REGIONS ON A THREE-DIMENSIONAL MODEL OF LTRA

Figure 3.6(a) and (b) show two views of a three-dimensional model of LtrA constructed by threading the LtrA sequence on the X-ray crystal structure of HIV-1 RT heterodimers (Blocker et al., 2005). In this model, LtrA is assumed to be a dimer (Saldanha et al., 1999; Rambo & Doudna, 2004) with one subunit (monomer A) having a structure analogous to the catalytic p66 subunit of HIV-1 RT and the second subunit

(monomer B) modeled based on the p51 subunit. The latter is an equivocal but necessary assumption since HIV-1 RT is the only available RT dimer structure (Figure 3.6(a) and (b); Blocker et al., 2005).

Regions A, B, E, and H, conserved for efficient splicing of the wild-type intron but not for residual splicing of the ΔDIVa intron in the unigenic evolution analysis (Figure 3.4), are potentially involved in DIVa binding. In Figure 3.6(c) and (d), these regions are highlighted in red in the three-dimensional model of LtrA. Strikingly, regions A, B, and E of monomer A lie in proximity to each other in the extended N-terminal fingers domain, while part of region H is on the opposed surface of the monomer A thumb, potentially forming a binding pocket for DIVa.

Regions C, F, and J are required for splicing both of the wild-type and ΔDIVa introns and are potentially involved in binding conserved catalytic core regions and/or play essential roles in maintaining protein structure. These regions are colored purple in the model in Figure 3.6(c) and (d). Regions C and F comprise major parts of the palm, while region J is in the thumb, facing away from the fingers.

The remaining regions (D, G, I, and K; colored in blue in 3.6(c) and (d)) are more constrained in variants that splice the ΔDIVa intron than those that splice the wild-type intron. Region D is in palm. Regions G and I are major components of the thumb, whereas region K is an extension from the thumb into the DNA-binding domain. These regions are also potentially involved in interactions with the intron's catalytic core.

3.3 ANALYSIS OF SITE-DIRECTED MUTANTS

To identify amino acid residues potentially involved in binding DIVa, I compared unigenic evolution data for LtrA variants that splice the wild-type and ΔDIVa introns to identify charged residues, particularly in regions A, B, E, and H, that were conserved for efficient splicing of the wild-type intron but not for residual splicing of the ΔDIVa intron.

Alanine substitutions of single or multiple positions were then constructed in both the pELG2-ΔORF and pELG2-ΔORF/ΔDIVa backgrounds to test their effect on splicing of the wild-type and ΔDIVa intron, respectively. A total of 17 such mutants, along with N-terminal deletion mutants ΔN10 and ΔN20, which were analyzed biochemically in Chapter 2, were compared for their ability to splice the wild-type and ΔDIVa introns *in vivo* by the FACS assay and RT-PCR, respectively. The results are summarized in Table 3.1. ΔN10 and ΔN20 were included because the previous biochemical analysis showed that they abolish high-affinity binding to the intron RNA, as expected for impaired interaction with DIVa (chapter 2; Cui et al., 2004).

Mutations that specifically affect the binding of DIVa are expected to reduce the splicing efficiency for the wild-type intron, but have little or no effect on the residual splicing of the ΔDIVa intron. The following eight mutants satisfied these criteria: ΔN10, ΔN20, R137A, KKS182,185,187→A (referred to as KKS), YKF191-193→A (referred to as YKF), Y204A, KKD407-409→A (referred to as KKD), DLE425-427→A (referred to as DLE) (Figure 3.7). These eight mutants were then subjected to Western analysis to assess whether the splicing defects for the wild-type intron might reflect decreased amounts of soluble and total LtrA protein (Figure 3.8(a)-(d)).

ΔN10, ΔN20, and a triple mutant YKF gave no detectable splicing of the wild-type intron as judged by GFP expression in the FACS assay (Figure 3.7(a)) but were indistinguishable from the wild-type protein in their ability to splice the ΔDIVa intron, as judged by RT-PCR (Figure 3.7(b)). All three mutants gave less soluble protein than wild type in both the pELG2-ΔORF and pELG2-ΔORF/ΔDIVa backgrounds. The expression levels of ΔN20 and YKF were lowest among all the mutants, about 5-10% of wild type, whereas ΔN10 was about 20-25% of wild type (Figure 3.8). In the structural model, N10 and N20 are at the N-terminus, presumably structurally flexible. YKF in monomer A is

located within insertion 3a on the surface of the fingers region and may directly affect the interaction with DIVa. Interestingly, in monomer B, YKF is mostly at the dimer surface with a small part on the surface of the bottom of the predicted binding pocket.

Two point mutations, R137A, and Y204A, decreased the splicing efficiency of the wild-type intron by 70-80% in the FACS assay (Figure 3.7(a)) but spliced the ΔDIVa intron at almost wild-type efficiency in the RT-PCR assay (Figure 3.7(b)). Both mutants gave high levels of soluble protein in the pELG2-ΔORF background, but about half of the wild-type level in pELG2-ΔORF/ΔDIVa, even though the splicing efficiency with the ΔDIVa intron was similar to that of the wild-type protein (Figure 3.8). In the structural model, R137 in monomer A is located within insertion 2a on the surface of the palm directly below the putative binding pocket for DIVa and is a surface residue in monomer B. Y204, which is located within RT-4 just downstream of insertion 3a, is on the surface of the fingers in monomer A and partially buried at the dimer interface in monomer B (Figure 3.5(e) and (f)). Thus both mutations could affect the interaction with DIVa directly.

The remaining mutants that showed the desired phenotypes, KKS, KKD, and DLE, inhibited splicing of the wild-type intron by 60-80% in the FACS assay (Figure 3.7(a)), but had no detectable effect on ΔDIVa splicing activity in the RT-PCR assay (Figure 3.7(b)). KKS gave wild-type level of soluble protein in the pELG2-ΔORF background but only about half of the wild-type level in pELG2-ΔORF/ΔDIVa, while KKD and DLE gave wild-type level of soluble proteins in both backgrounds (Figure 3.8). KKS is located within insertion 3a and on the surface of monomer A in the structural model. However, all three of the residues are at the dimer surface in monomer B. KKD and DLE had 30-40% wild-type splicing in the FACS assay. KKD is in insertion ti, and DLE is located immediately down stream of the insertion. In monomer A of the model,

these residues are on the surface of the thumb, facing away from the prospective DIVa binding pocket. It is possible that the orientation of ti is incorrect in the model, or that these mutations affect binding of DIVa indirectly by affecting the structure of ti. In monomer B, residues in both KKD and DLE are also on the protein surface.

It is worth noting that several mutants (YKF, KKS, Y204A) involve residues that are at the dimer interface in at least one monomer in the structural model, raising the possibility that, in addition to directly interact with DIVa, these positions could indirectly influence the LtrA/intron interaction by affecting dimerization of LtrA.

All six of the mutants identified above are excellent candidates for biochemical analysis to directly assess their effect on DIVa binding.

3.4 DISCUSSION

In this chapter, I screened an LtrA library for functional variants that spliced the Ll.LtrB intron in the absence of the high-affinity protein-binding site, DIVa. One hundred seventy mutants that remained active for splicing the ΔDIVa intron contained 575 missense and 370 silent mutations. I used unigenic evolution analysis to identify regions conserved in LtrA variants that splice the ΔDIVa intron. A superimposition of these data with those for variants that splice the wild-type intron reveals three categories of regions: regions required for efficient splicing of the wild-type intron but not for residual splicing of the ΔDIVa intron; regions required for splicing both the wild-type and ΔDIVa introns; and regions that are more conserved in variants that splice the ΔDIVa intron than those that splice the wild-type intron (Figure 3.4).

3.4.1 Three categories of regions

Potential DIVa-binding regions are those conserved for the efficient splicing of the wild-type intron splicing but not for residual splicing of the ΔDIVa intron. Four such

regions were identified in the unigenic evolution analysis: regions A, B, and E in the RT domain and H in domain X. In Figure 3.5(c) and (d), these regions are shown in red in a structural model of LtrA based on an X-ray crystal structure of HIV-1 RT dimer. In the model, regions A, B, and E of monomer A are located next to each other in the extended fingers domain, while region H is located in the thumb domain partially facing the fingers, forming a potential binding pocket for DIVa. Figure 3.5(c) and (d) also displays the positions of insertions 3a and ti, which are located in regions E and H, respectively. These two insertions are frequent sites of mutations that gave the expected phenotype of inhibiting efficient splicing of the wild-type intron while preserving residual splicing of the ΔDIVa intron (see next section).

Regions C, F, and J (colored in purple in Figure 3.5 (c) and (d)) were required for splicing both the wild-type and ΔDIVa introns. These regions span most of the palm domain, with a small extension to the base of the thumb (region J). They may bind the intron catalytic core and/or be required to maintain the core structure of the protein.

Finally, four other regions (D, G, I, and K, colored in blue in Figure 3.5 (c) and (d)) were conserved only for splicing the ΔDIVa intron but not for efficient splicing of the wild-type intron. The greater constraints on these four regions for splicing of the ΔDIVa intron may indicate that some weaker interactions with the catalytic core region of the RNA assume greater importance when DIVa is deleted. It is also possible that in the absence of DIVa, additional regions on the proteins surface are exposed to bind the catalytic core, or that RNA structural changes resulting from deletion of DIVa affect protein binding.

3.4.2 Point mutations analyzed for identifying direct DIVa binding

To support the involvement of regions A, B, E, and H in DIVa binding, it is important to identify residues in these regions that directly interact with DIVa. Based on

the unigenic evolution analysis, I chose a series of positions mainly from the above regions for site-directed mutagenesis. Seventeen alanine-substitution mutants were constructed containing one to three clustered mutations per protein. All the mutants were first tested for their ability to splice the wild-type intron in pELG2-ΔORF in the FACS assay, and those that showed decreased splicing of the wild-type intron were then tested for their ability to promote residual splicing of the ΔDIVa intron in pELG2-ΔORF/ΔDIVa by RT-PCR. Western analysis was carried out to assess whether impaired splicing of the wild-type intron might reflect decreased levels of soluble protein. Eight mutants, including ΔN10 and ΔN20, which were analyzed biochemically in chapter 2, showed the expected *in vivo* splicing phenotypes indicating their possible involvement in interaction with DIVa (see Table 3.1).

The biochemical analysis of the ΔN10 and ΔN20 in chapter 2 showed that both mutants have dramatically decreased binding affinity for the wild-type intron, as expected for mutations that impair the binding of DIVa, but remain capable of splicing the wild-type intron efficiently at the high protein and RNA concentrations used in *in vitro* assays (Cui et al., 2004). *In vivo*, both mutants gave greatly decreased splicing of the wild-type intron in the FACS assay (Figure 3.7(a)), while retaining the ability to promote splicing of the ΔDIVa intron in the RT-PCR assay (Figure 3.7(b)). The greatly decreased splicing efficiency for the wild-type intron *in vivo* presumably reflects that the concentrations of both intron and protein are much lower than those in *in vitro* assays, amplifying the importance of high-affinity binding to DIVa. The Western analysis revealed that both the ΔN10 and ΔN20 mutants had reduced amounts of soluble protein in either the pELG2-ΔORF or pELG2-ΔORF/ΔDIVa backgrounds. In chapter 2, I demonstrated that the LtrA protein tends to be unstable *in vivo* when it cannot stably bind to the intron RNA (Cui et al., 2004). Thus, the relatively low level of ΔN10 and ΔN20 expression may result at least

in part from the weaker binding of these mutant proteins to the intron RNA. The structural effects of the N-terminal deletions could also decrease protein stability.

All six alanine substitution mutants (R137A, KKS, YKF, Y204A, KKD, and DLE) involve surface residues in monomer A in the LtrA structural model, surrounding the putative DIVa binding pocket. Potentially, all these positions could interact directly to DIVa. Except for R137, all the mutated residues are located in either insertion 3a or ti, suggesting that these insertions may play a critical role in DIVa binding. R137 is in insertion 2a, located at the bottom of the pocket.

In addition, in monomer B, several mutants (KKS, YKF, and Y204A) are at the dimer surface in the structural model, indicating a possible requirement of dimerization for DIVa binding but not for binding the catalytic core. It is also possible that these mutations affect the structure of another region of the protein involved in binding DIVa. Rambo & Doudna (2004) reported that the wild-type intron, DIVa, and DI, all bound the LtrA protein with a stoichiometric RNA:protein ratio of 1:2, and only upon binding of one of the above RNAs did the protein form a stable dimer. However, their results do not indicate whether dimerization is necessary for binding the cognate RNA species.

In summary, the unigenic evolution analysis and site-directed mutations combined with three-dimensional structural modeling support the hypothesis that the extended N-terminal finger region of LtrA is involved in high-affinity binding of DIVa, possibly forming a DIVa-binding pocket in combination with parts of the thumb domain, including insertion ti. Other regions of the RT and thumb (X domains) are potentially involved in binding the intron's conserved catalytic core regions to promote formation of the active RNA structure.

3.5 SUMMARY

Mobile group II introns encode RTs that bind specifically to the intron RNA to stabilize the catalytically active RNA structure for RNA splicing and reverse splicing. Previous studies with the *L. lactis* Ll.LtrB intron suggested a model in which the intron-encoded protein, LtrA, binds first to a high-affinity binding site in DIVa, an idiosyncratic stem-loop structure at the beginning of the LtrA ORF, and then makes weaker but critical secondary contacts with the conserved catalytic core region. In the absence of DIVa, group II intron-encoded proteins can promote varying degrees of residual splicing by binding directly to the catalytic core. Here, we used an *E. coli* genetic assay that permits the rapid screening of LtrA mutant libraries in conjunction with unigenic evolution analysis to identify LtrA regions potentially required for interaction with different regions of the intron RNA. Regions required for efficient splicing of the wild-type intron, but not for residual splicing of a ΔDIVa intron, are potentially involved in DIVa binding and include the N-terminus, RT-1, RT-3a and a small insertion (ti) in the thumb/domain X region. In a three-dimensional model of LtrA, these regions comprise a potential RNA-binding pocket that includes the extended N-terminal fingers domain and an opposed surface of the thumb. Regions required for splicing both the wild-type and ΔDIVa introns potentially interact with critical catalytic core regions and include an extended binding surface on the thumb and palm. Four regions, which include thumb and some surface residues on fingers, are more constrained in variants that splice the ΔDIVa intron but not the wild-type intron, potentially reflecting greater dependence on other catalytic core contacts when binding to DIVa is impaired. Analysis of site-directed mutants supports the hypothesis that the entended N-terminal fingers region and possibly part of the thumb contribute to DIVa binding.

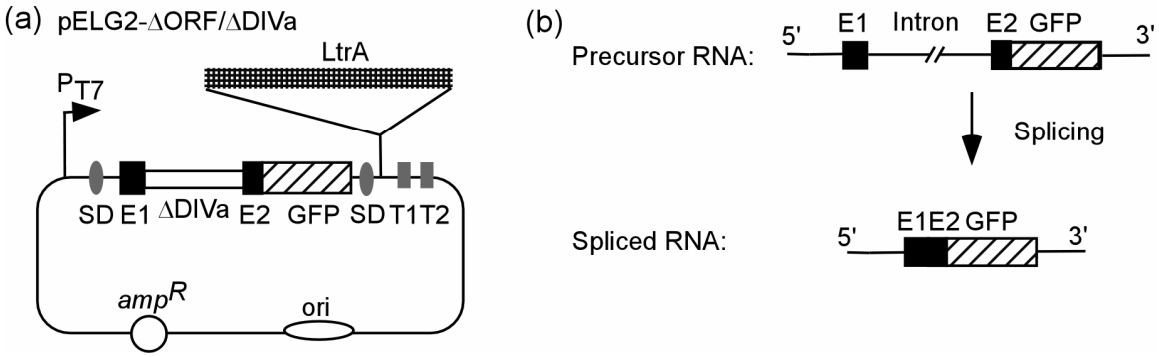


Figure 3.1. *E. coli* genetic assay of LtrA-promoted splicing of the ΔDIVa intron.

(a) Map of plasmid pELG2-ΔORF/ΔDIVa. The plasmid contains an *ltrB*/GFP fusion cloned downstream of a T7lac promoter (arrow) and ϕ10 Shine-Dalgarno sequence (SD). The *ltrB*/GFP fusion is comprised of the first eight codons of the ϕ10 gene fused to a 5'-exon segment (E1) of the *L. lactis ltrB* gene, the Ll.LtrB-ΔORF/ΔDIVa intron (open rectangle), and a 3'-exon segment (E2), with E2 linked in-frame to the second codon of GFP (hatched). The LtrA ORF is cloned downstream of the GFP coding sequence. T1 and T2 are *E. coli* rrnB transcription terminators. (b) Schematic showing splicing of the ΔDIVa intron linked to the expression of GFP. Splicing leads to the expression of GFP with a short N-terminal extension (42 amino acid residues) corresponding to the ϕ10 sequence and ligated exons. In the absence of splicing, translation is terminated by stop codons within the intron.

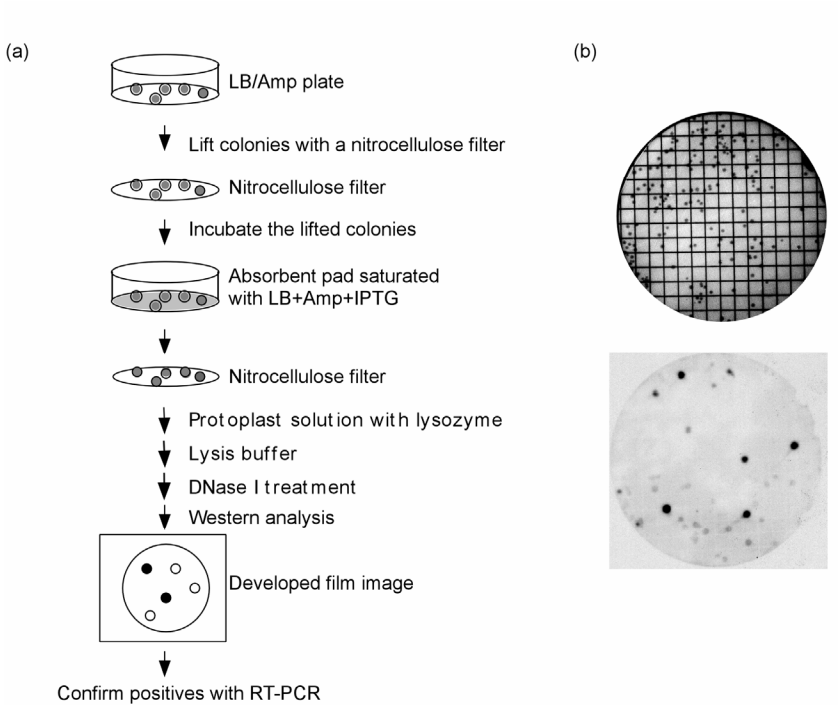


Figure 3.2 Colony-based immunoblot assay for GFP expression.

(a) Flow chart of the assay. Colonies growing on plates containing LB medium plus ampicillin were duplicated onto nitrocellulose filters and then incubated on an absorbent pad saturated with LB medium containing ampicillin and IPTG for 4 h at 37°C to induce plasmid expression. The colonies were then lysed and subjected to immunoblot analysis, as described in Materials and Methods. (b) Scanned filters. *E. coli* cells transformed with pELG2-ΔORF/ΔDIVa and with pELG1-ΔORF/ΔDIVa (splicing competent and splicing defective, respectively) were mixed at a ratio of 1:10, then plated on LB medium plus ampicillin and subjected to colony based immunoblot assay as described in (a). The top shows an optical scan of the filter before treatment, and the bottom shows a chemiluminescence exposure of the filter after immunoblotting. Splicing-competent colonies producing GFP appear black against a light background.

Figure 3.3 Summary of mutations in the 170 LtrA variants that could promote splicing of the ΔDIVa intron.

A library of random LtrA mutants was constructed in pELG2-ΔORF/ΔDIVa background and screened by using the colony immunoblot assay to identify splicing-competent variants. Positive colonies were confirmed to have splicing activity by RT-PCR and sequenced. Amino acid substitutions are listed below the wild-type sequence, with the frequency of occurrence indicated as following: plain, 1X; bold, 2X; and underlined, ≥ 3X. Numbers represent occurrence of silent mutations. Conserved regions RT-0 to RT-7, insertions RT-2a, -3a, -4a, and -7a, and domains X, D, and En are delineated above the sequence. Insertion ti is marked with a thick line.

Figure 3.3

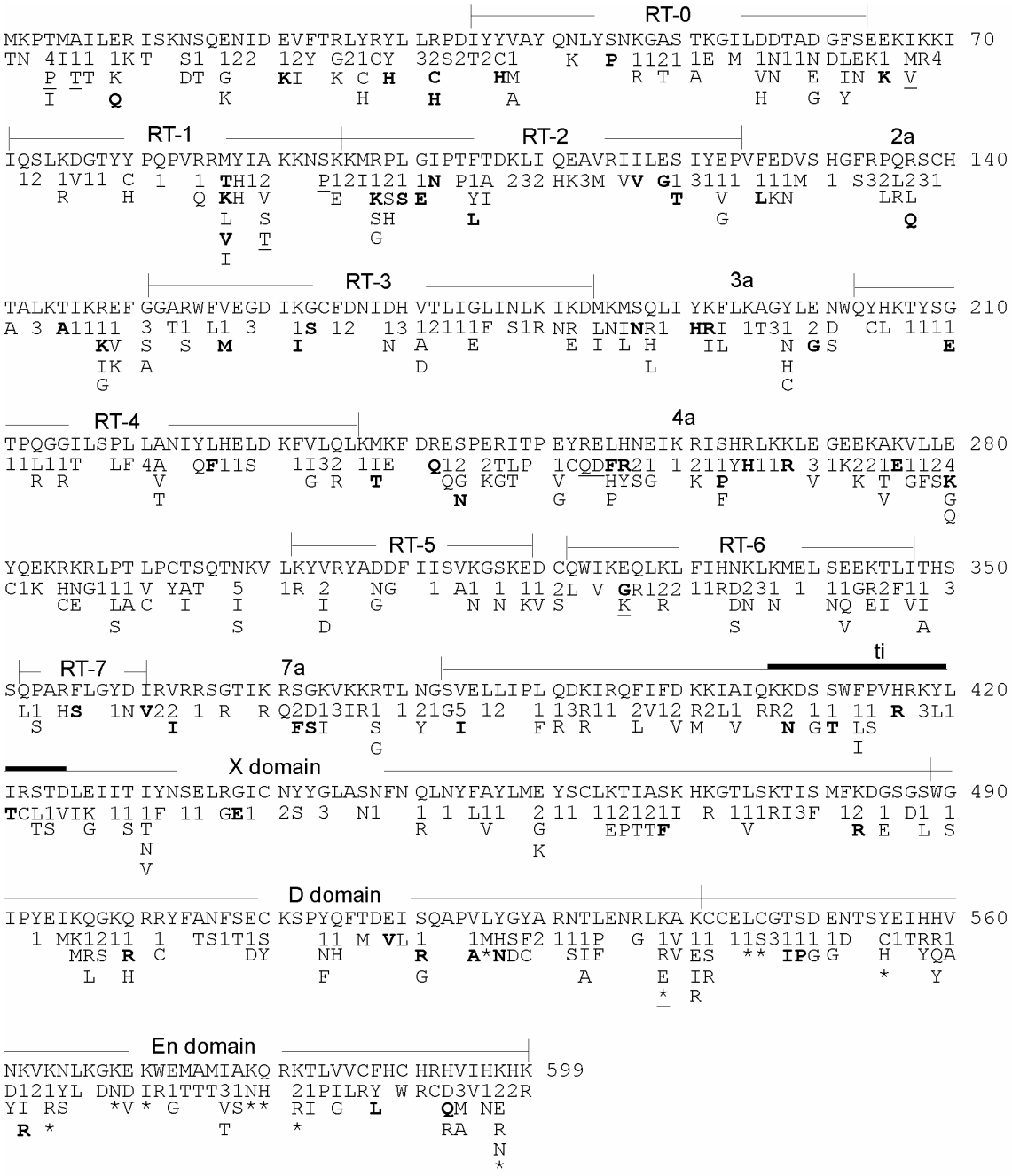
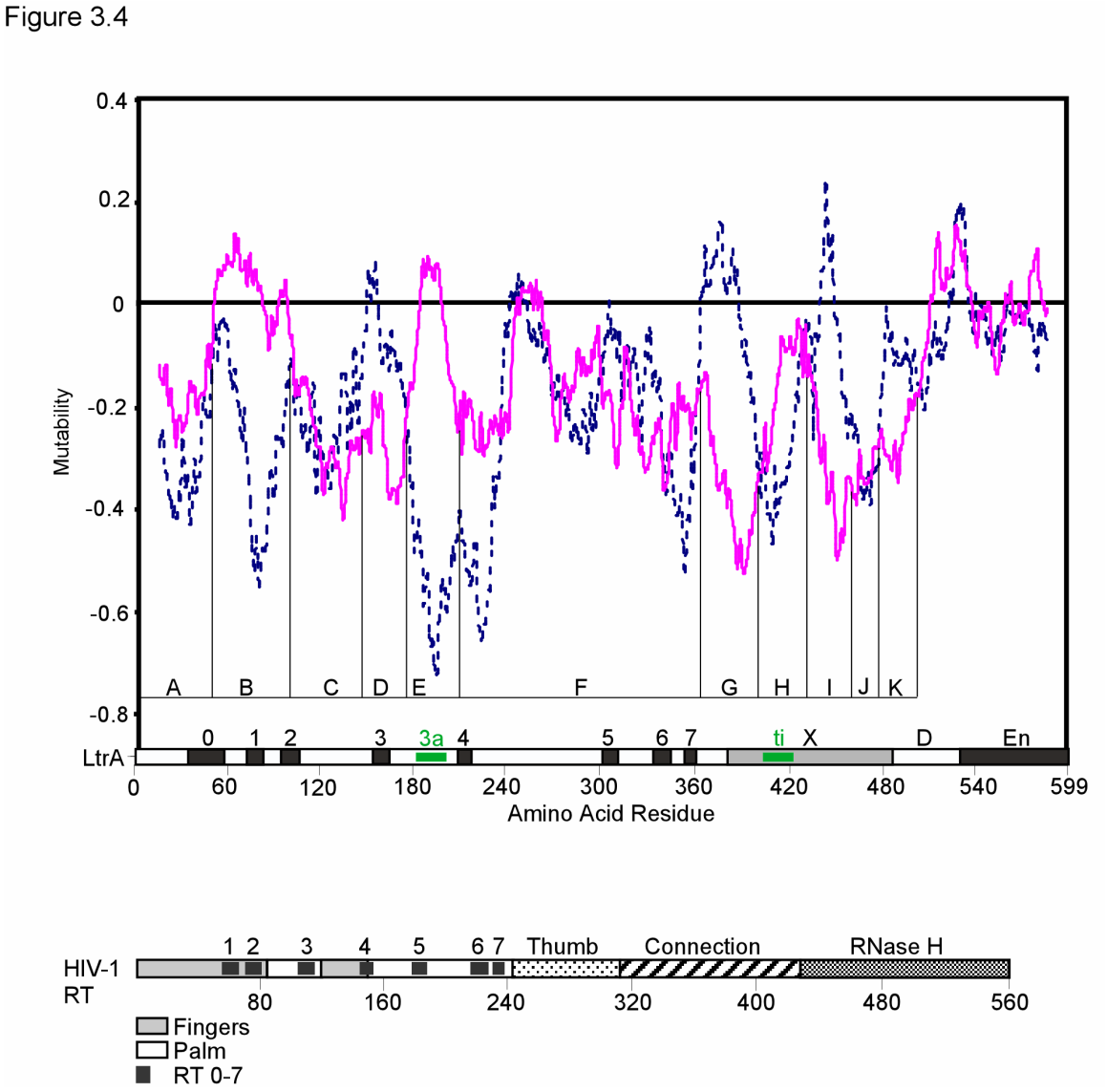

RT-0
RT-1
RT-2
2a
RT-3
3a
RT-4
4a
RT-5
RT-6
RT-7
7a
ti
X domain
D domain
En domain

Figure 3.4 Mutability plot based on unigenic evolution analysis of LtrA variants that splice the ΔDIVa intron.

The mutability value (M) was calculated for a 25 amino acid residue-sliding window, as described in Materials and Methods, and mutability values were plotted against the amino acid at the center of the window. Negative values indicate hypomutability, with a maximum value of -1 indicating no missense mutations, and positive values indicate hypermutability. A region is considered significantly hypomutable if it has a p-value <0.05. A previous plot for LtrA variants that efficiently splice the wild-type intron is shown in dashed lines for comparison. The locations of conserved sequence motifs RT0-7 and LtrA domain boundaries are shown on the x-axis. Insertions RT-3a and ti are indicated in green.

Figure 3.4


0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
Mutability
A
B
C
D
E
F
G
H
I
J
K
0
1
2
3
3a
4
5
6
7
ti
X
D
En
LtrA
0
60
120
180
240
300
360
420
480
540
599
Amino Acid Residue
HIV-1
RT
1 2
3
4
5
6 7
Thumb
Connection
RNase H
80
160
240
320
400
480
560
Fingers
Palm
RT 0-7

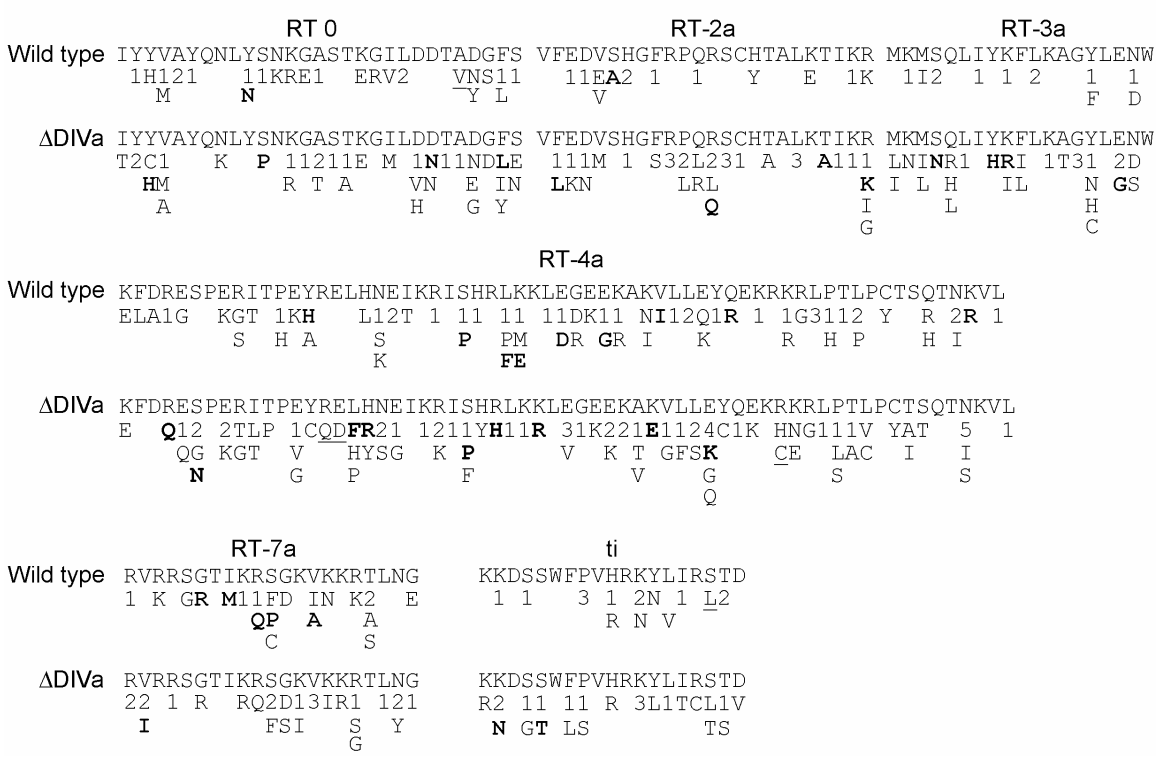


Figure 3.5 Comparison of mutations in insertion regions in LtrA variants that splice the wild-type and ΔDIVa introns.

As in Figure 3.3, amino acid substitutions are listed below the wild-type sequence, with the frequency of occurrence indicated as following: plain, 1X; bold, 2X; and underlined, ≥ 3X, and numbers represent the occurrence of silent mutations.

Figure 3.6 Three-dimensional model of LtrA.

(a) and (b) are two views of the 3-D model of LtrA with monomer A in white and monomer B in cyan. Regions of interest are labeled as in Blocker et al. (2005). (c) and (d) are two views of LtrA with colored regions to differentiate conservation in LtrA variants that promote splicing of the wild-type and ΔDIVa introns. Red regions are conserved for efficient splicing of the wild-type intron, but not for splicing of the ΔDIVa intron. Blue regions are more conserved for splicing the ΔDIVa intron than the wild-type intron. Purple regions are conserved for splicing both the wild-type and ΔDIVa introns. Approximate positions of regions A-K from Figure 3.4 are labeled. Insertions RT-3a and ti are also labeled. (e) and (f) are two views of LtrA with monomer A in white and monomer B in cyan. Surface residues are in orange, and residues at the dimer interface are in blue). The figure and modeling were done by Dr. Forrest Blocker.

Figure 3.6

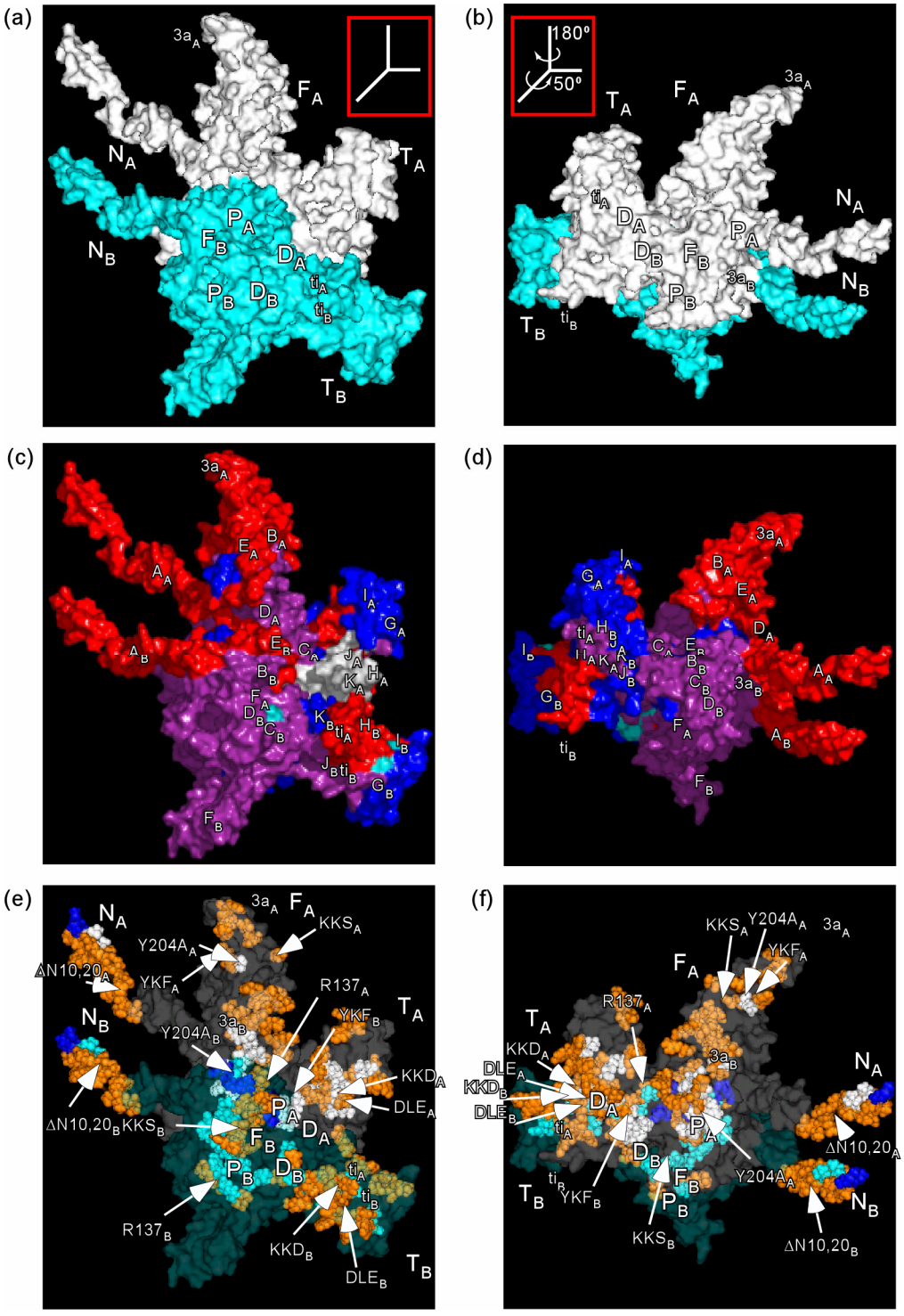


Table 3.1 Site-directed mutants of LtrA.

Mutation	Region [a]	Domain	WT [b,c] splicing	ΔDIVa [b,c] splicing	Protein level in WT [b,c]	Protein level in ΔDIVa [b,c]
ΔN10	A	N-ter	-	++++	++	++
ΔN20	A	N-ter	-	++++	+	+
*D56A	B	RT0	++++	nt	nt	nt
*S63A	B	RT0	++++	nt	nt	nt
KD75,76→A	B	RT1	++++	nt	nt	nt
*R85A	B	RT1	++++	nt	nt	nt
*R137A	C	RT2a	+	++++	++++	+++
*K182A	E	RT3a	++++	nt	nt	nt
*K185A	E	RT3a	++++	nt	nt	nt
*S187A	E	RT3a	++++	nt	nt	nt
*KKS	E	RT3a	+	++++	++	++
*K192A	E	RT3a	++++	nt	nt	nt
*YKF	E	RT3a	-	+++	+	+
*E200A	E	RT3a	+++	++++	nt	nt
*Y204A	E	RT4	+	++++	+++	+++
KK407,408→A	H	ti	++++	nt	nt	nt
KKD	H	ti	++	++++	++++	++++
*R422A	H	ti	+++	nt	nt	nt
DLE	H	ti	++	++++	+++	++++

[a]Regions are defined by unigenic evolution analysis as in Figure 3.4. [b]WT refers to the wild-type intron, and ΔDIVa refers to the ΔDIVa intron. [c]Splicing activity and protein expression levels were expressed by ++++, wild-type level; +++, ≥ 50% wild-type level; ++, ≥ 30% wild-type level; +, detectable; and -, undistinguishable from untransformed cells; nt, not tested. [d]KKS: mutant KKS182,185,187→A; YKF: mutant YKF191-193→A; KKD: KKD407-409→A; DLE: mutant DLE425-427→A. Mutants with the phenotype expected for inhibition of DIVa binding are shaded in grey. Constructs made by Ms. Qin Wang are with an asterisk.

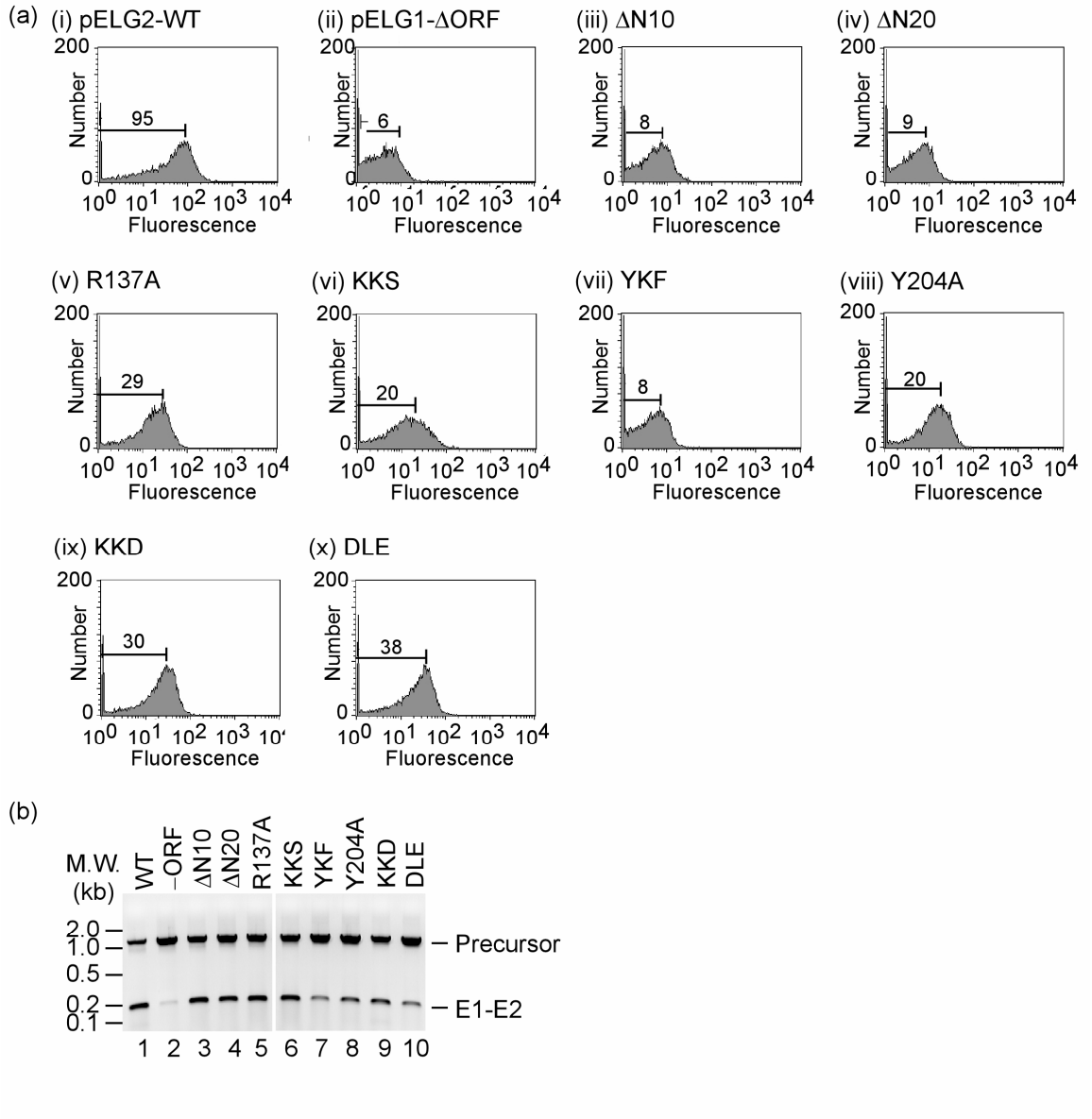


Figure 3.7 FACS and RT-PCR analysis of *in vivo* splicing with LtrA variants that are unable to efficiently splice the wild-type intron, but can promote residual splicing of the ΔDIVa intron.

(a) FACS assay of *in vivo* splicing of the wild-type intron. (b) RT-PCR for *in vivo* splicing of the ΔDIVa intron.

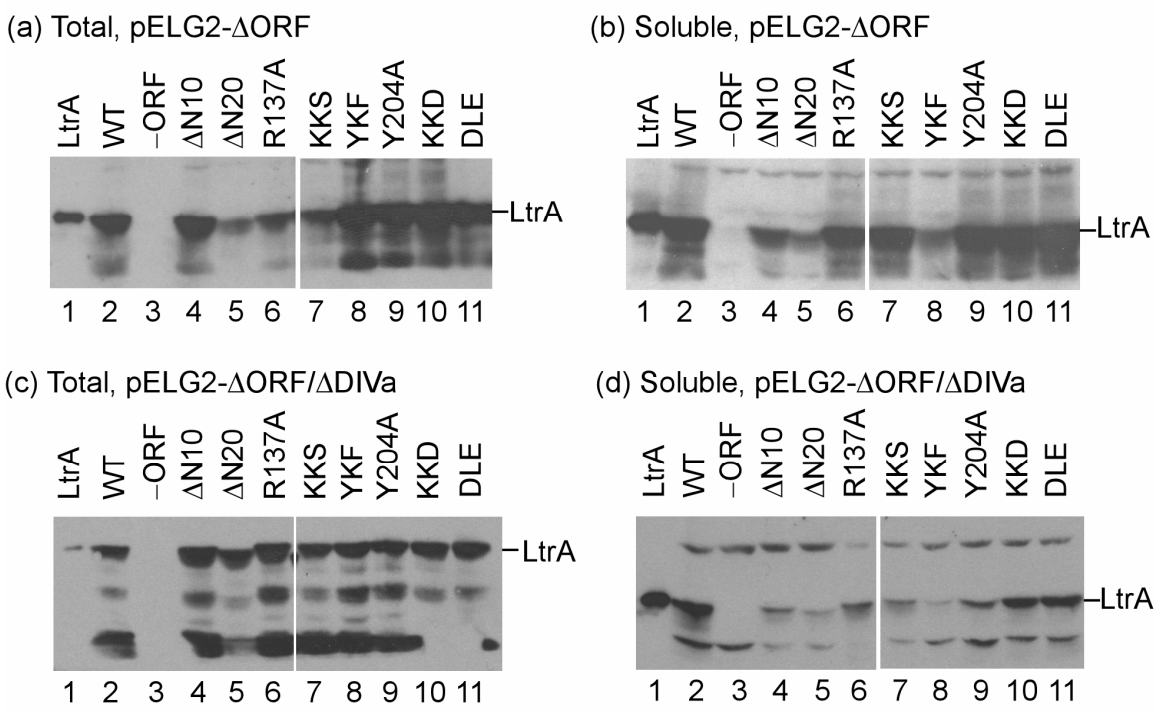


Figure 3.8 Immunoblot analysis of LtrA proteins expressed from pELG2-ΔORF and pELG2-ΔORF/ΔDIVa.

(a) and (b) Total and soluble protein fractions from cells expressing the indicated LtrA proteins from pELG2-ΔORF. (c) and (d) Total and soluble protein fractions from cells expressing the indicated LtrA proteins from pELG2-ΔORF/ΔDIVa. Samples were loaded based on $O.D._{595}$ measurement of cell density. Each transferred membrane was stained with AuroDye after blotting to confirm even loading. Proteins were separated in an 8% polyacrylamide/1% SDS gel, blotted to a nitrocellulose membrane, and probed with polyclonal anti-LtrA antibody (from Gary Dunny, U. Minnesota). Smaller bands in (a) and (c) are likely degradation products of LtrA. In (b) and (d), one larger band and one smaller band result from non-specific binding by the antibody, as they are also present in ΔORF lanes, which have no LtrA expresson.

Chapter 4: Expression of the Ll.LtrB intron RNA and IEP in mammalian cells

The mobile group II intron Ll.LtrB now works extremely well for gene targeting in bacteria. Target site recognition requires base-pairing between the EBS and δ sequences of the intron RNA with the target DNA sequences, as well as interactions between the IEP (LtrA) and DNA sequences flanking the base-paired region (Guo et al., 2000; Mohr et al., 2000). Guo et al. (2000) developed a genetic method to select modified introns that insert into essentially any plasmid-borne target gene from a combinatorial group II intron library. Karberg et al. (2001) demonstrated that bacterial chromosomes could be targeted by the Ll.LtrB intron, and Zhong et al. (2003) developed highly efficient methods for selecting integration events by incorporating selectable markers into DIV of the intron, which is also valuable to make genome-wide knockouts.

The intron RNA inserts into a target site in either the same or the opposite direction relative to that of target gene transcription. When it inserts into a coding sequence in the sense-strand orientation, the intron will be transcribed and could potentially be spliced out from the precursor RNA if and only if the IEP is supplied, resulting in a conditional disruption. Frazier et al. (2003) used a modified Ll.LtrB intron to deliver a phage-resistance gene (*abiD*) and an antibiotic-resistance gene (*tetM*) separately into the sense strand of the chromosomal *mleS* gene in *Lactococcus lactis*, and showed that it was possible to control the expression of *mleS* gene by regulating the expression of the LtrA protein. Generally, the use of group II introns to obtain conditional disruptions is limited by whether or not the sequences flanking the intron-insertion sites can support efficient splicing (Karberg et al., 2001).

Group II introns are minimally dependent on host-specific factors for mobility, and derivatives of the Ll.LtrB intron have been shown to work efficiently for gene targeting in a variety of Gram-negative and Gram-positive bacteria, including *E. coli*, *L. lactis*, *Salmonella typhimurium*, *Shigella flexneri*, and *Staphylococcus aureus* (Karberg et al., 2001; Zhong et al., 2003; Fraizer et al., 2003; and Zhong, Novick, & Lambowitz, unpublished). The general strategy to obtain retargeted group II introns (introns specifically targeting genes other than the wild-type target site) is to identify the best IEP-recognition sequences and then modify the intron EBS1, EBS2, and δ sequences to base pair with the IBS1, IBS2, and δ' sequences in the DNA target site (Guo et al., 2000; Figure 1.1). Retargeted Ll.LtrB introns are now routinely designed by using a computer algorithm based on databases of efficiently targeted introns and their target sites (Perutka et al., 2004). This algorithm predicts multiple, rank-ordered target sites in virtually any gene and designes primer sequences for modifying the intron to insert efficiently into those sites.

The ability to use the Ll.LtrB intron for efficient gene targeting is due largely to the development of an efficient *E. coli* expression system for this intron, which enabled extensive genetic and biochemical studies that characterized protein/RNA interaction, splicing, and targeting rules. It is now of great interest to extend these methods to manipulate the more complex genomes and gene expression of higher organisms. Methods currently available for these purposes in higher organisms include homologous recombination (Thomas & Capecchi, 1987), antisense RNA (Green et al., 1986), and RNA interference (Fire et al., 1998). Homologous recombination has been widely used for generating gene knockouts. However, it is labor intensive and inefficient. Both antisense and siRNAs produce "knockdowns", instead of "knockouts", and are unstable, likely generating only transient effects. Once the input RNA molecules are degraded, the

expression of down-regulated genes may recover. Or plasmids expressing siRNAs could be integrated into the genome for lasting effects, however, carrying the risks resulting from random integrations. Group II introns insert in a highly site-specific manner into genes and disrupt them permanently or conditionally, controllable by the IEP expression (Karberg et al., 2001; Fraizer et al., 2003). Group II introns can also be used to generate site-specific double-strand breaks (DSB), which stimulate homologous recombination, enabling the introduction of smaller modifications, such as point mutations (Karberg et al., 2001). Homing endonucleases also cleave DNA sequence-specifically and generate DSBs. However, unlike group II introns, changing the specificity of homing endonucleases requires protein engineering of the enzymes (Epinat et al., 2003). In addition, group II introns are functional as vectors to deliver a desired gene to a specific location (Fraizer et al., 2003). If group II introns could be made to work efficiently in higher organisms, their impact on functional genomics and gene therapy would be profound.

Guo et al. (2000) designed and selected group II introns that inserted efficiently into the HIV-1 provirus DNA and the human gene encoding the HIV-1 co-receptor CCR5 in bacterial plasmid assays and showed that purified RNPs, when introduced via liposome-mediated transfection, retain some activity in human cells for insertion into a plasmid target site. However, the activity was very low, requiring nested PCR for detection. In part, this low efficiency reflects inefficient introduction of RNPs by the transfection procedure. In this chapter, I describe the development of a system to express the Ll.LtrB intron RNA and IEP, LtrA, in cultured mammalian cells. The advantages of this approach are that DNA constructs for expressing the intron RNA and IEP can be easily and efficiently transfected into mammalian cells and potentially produce a

continuous supply of RNP particles for a few days, which can be especially critical if targeting can only occur during certain cell cycle or growth phases.

To express the LtrA protein in mammalian cells, I found it necessary to optimize codon usage of the LtrA ORF to that preferred in human cells. Here, I show that the LtrA protein expressed in this manner is abundant and active in mammalian cells, as judged by detection of Ll.LtrB intron-specific RT activity in nuclear lysate and by IEP-dependent splicing of the Ll.LtrB intron *in vivo*. Further, when a nuclear localization signal is fused to its C-terminus, the LtrA protein is correctly localized to the nucleus, as required for chromosomal targeting. I also established stable cell lines that express active LtrA protein. The Ll.LtrB intron RNA could be expressed using a Pol I, Pol II, or Pol III promoter. However, intron RNA transcribed from a Pol I promoter spliced most efficiently in the presence of expressed LtrA protein *in vivo* Interestingly, when cells were provided with high concentrations of Mg^{2+} in the medium, the intron RNA could splice to some extent in the absence of the LtrA protein. Finally, I describe initial attempts to use the expression system for gene targeting.

4.1 EXPRESSION OF THE LL.LTRB IEP, LTRA IN TISSUE CULTURE

To express the LtrA protein in mammalian cells, I first cloned the LtrA ORF downstream of a human cytomegalovirus (CMV) immediate early promoter in the plasmid pCMV/myc/nuc (Invitrogen) to generate pCMV-LtrA (Figure 4.1(a); see also Materials and methods). In this construct, three tandem nuclear localization sequences (NLS) from simian virus 40 (SV40) and a c-myc epitope were fused to the C-terminus of the LtrA ORF**.** The CMV promoter is a constitutive promoter recognized by RNA polymerase II and leads to high-level expression of its downstream sequences (Spaete & Mocarski, 1985). Detecting no protein expression by using Western analysis with either an anti-LtrA antibody (Gary Dunny, U. of Minnesota; not shown) or an anti-myc

antibody (Invitrogen), I tried to improve LtrA expression by inserting a 133-bp intervening sequence (IVS, from vector pIRES, Clontech) immediately following the CMV promoter (pCMV-IVS, figure 4.1(b)), resulting in pCMV-IVS-LtrA (Figure 4.1(c)). The IVS encodes a nuclear intron that is efficiently spliced following transcription, which could increase protein production (Le Hir et al., 2003). Again, when transfected with pCMV-IVS-LtrA, no LtrA expression was detected in HEK 293, HeLa, or COS-7 cells by using Western analysis (Figure 4.4, lane 3 and not shown).

4.1.1 Codon optimization of the LtrA ORF

I noticed that most codons of the LtrA ORF are highly unfavorable for mammalian cells, and suspected that this might contribute to its inefficient translation. Therefore, I modified the LtrA coding sequence by changing every codon to the most favored one in higher eukaryotes (Figure 4.2) based on reported codon usages (Haas et al., 1996). The most common way of synthesizing a gene is to use overlapping primers as templates, amplify the entire sequence in several rounds of PCR, and then clone the sequence into an expression vector. A major problem with this strategy is that the longer the sequence being synthesized, the greater the chance for introducing mutations in each clone, during both primer synthesis and PCR. Fixing the mutations is time-consuming and costly. To avoid this problem, I divided the codon-optimized LtrA sequence, which included an SV40 nuclear localization signal (NLS) fused at the C-terminus, into six segments of about 300-bp in length, each of which was flanked by two restriction sites unique to the segment. Oligonucleotides of 98- to 120-base long were synthesized and used as overlapping templates, and short primers of 30-45 bases with common restriction sites (*Hind*III and *Xba*I) at their 5' ends were used to amplify each fragment by PCR (Figure 4.3, also see section 6.2). The PCR product of each fragment was then cloned into the pBluescriptKS vector individually for sequencing, except for those of the last two

segments at the C-terminus, which were PCR amplified into one piece, then cloned and sequenced. The complete LtrA sequence was obtained by ligating all the mutation-free segments after restriction enzyme digestions (see Figure 4.3, bottom panel). On average, only one out of five subclones sequenced was free of mutations. Thus, the chance of having an intact clone of the full-length ORF would be around one in 5^6, which is $1.5x10^4$, if the traditional strategy were used. On the contrary, I only needed to sequence about 40 subclones to obtain the correct codon-optimized LtrA clone, designated as pKS-hLtrA (Figure 4.1(d)), where hLtrA stands for the humanized LtrA gene.

4.1.2 Codon-optimized LtrA was abundantly expressed in human cells

The expression vector phLtrA was constructed by cloning the hLtrA ORF into the vector pIRES (Clontech; Figure 4.1(e)). The hLtrA gene is preceded by the immediate early CMV promoter and an IVS, as described above, and followed by a C-terminal fusion of the SV40 nuclear localization signal (NLS) and the SV40 poly A signal (pA). When transfected into HEK 293 cells, hLtrA was expressed well as shown in Western analysis in Figure 4.4a, lane 5, compared to no LtrA expression in untransfected cells (lane 2), cells transfected with vector pCMV-IVS (lane 3), or with pCMV-IVS-LtrA, encoding the non-codon-optimized LtrA ORF (lane 4). I also successfully obtained HEK 293 stable cell lines expressing hLtrA by selecting for neomycin-resistance clones after electroporating the cells with *Bam*HI-linearized phLtrA DNA (Figure 4.1(e)). However, the expression level in stable cell lines was always lower than that in transiently transfected cells (Figure 4.4(a), lanes 6 and 7). phLtrA and pCMV-IVS-LtrA have different vector backbones: pIRES and pCMV-IVS, respectively. In order to exclude any possible effects of vector backbone on protein expression, I replaced the hLtrA ORF with the non-optimized bacterial LtrA ORF in phLtrA to generate pLtrA (Figure 4.1(f)). pLtrA

did not give a detectable amount of LtrA protein (Figure 4.4(b), lane 2), just like pCMV-IVS-LtrA.

In Western analysis, a high-molecular weight band (>200 kDa) was present in variable amounts in some phLtrA samples (Figure 4.4(a), lane 5, about 1/3 of the amount of hLtrA and Figure 4.4(b), lane 4, less than 1/4 of the amount of hLtrA). When preparing the samples in Figure 4.4(b), I noticed that after being boiled in SDS-containing gel-loading buffer, each sample still had insoluble materials, which were pelleted with a two-minute centrifugation. The supernatant was loaded to lane 4, while the pellets were then dissolved in more loading buffer, boiled, and loaded in lane 5. Lane 5 does not show the large band, suggesting that it might be aggregated hLtrA protein, which is not dissociated when the amount of SDS in the loading buffer is limiting. Consistent with this hypothesis, the larger band was absent in phLtrA stable cell lines, where the hLtrA expression level was much lower (Figure 4.4(a), lanes 6 and 7).

4.1.3 Localization of hLtrA

To be able to target chromosomal DNA, intron RNPs must enter the nucleus. Thus, in the expression system, it is essential that the hLtrA protein is localized to the nucleus after translation. Using immunofluorescence, I found that LtrA expressed from transiently transfected phLtrA was localized to the nucleus in HEK 293 cells (Figure 4.5(a)) and COS-7 cells (Figure 4.5(b)). The LtrA protein expressed in the phLtrA stable cell lines was also localized to the nucleus (Figure 4.5(c)). By contrast, when COS-7 cells were transfected with construct phLtrAΔNLS (Figures 4.1(g)), LtrA was localized in the cytoplasm (Figure 4.5(d)), indicating that an NLS is necessary to direct the hLtrA protein to the nucleus. Interestingly, in transfected COS-7 cells, hLtrA expressed from phLtrA was localized to the nucleus but seemed to avoid nucleolar regions (Figure 4.5(e)), whereas in HEK 293 cells, the nuclear localization of hLtrA seemed uniform.

The intriguing non-nucleolar localization of hLtrA in COS-7 cells could also be an indication of protein aggregation, as suggested by Western analysis (last section). A difference between COS-7 cells and HEK293 cells is that the former constitutively expresses the SV40 T–antigen and supports replication of plasmids carrying the SV40 replication origin embedded in the SV40 promoter, as is the case for phLtrA. Because of phLtrA replication, more hLtrA protein was produced in COS-7 cells than in HEK 293 cells, indicated by stronger singnals in both immunofluorescence and Western analysis in COS-7 cells (not shown), and consequently more hLtrA aggregation occurred in COS-7 cells. It is possible that the aggregation occurs after the hLtrA protein enters the nucleus. The high density of nucleolar regions (Olson, 2004) may prevent large aggregates from spreading into those areas. Taken together, the localization and Western analysis data raise the possibility that hLtrA aggregation might occur in the cells under high expression conditions, which could potentially affect the intracellular localization and function in splicing and targeting.

Consistent with the aggregation hypothesis, our collaborator, Huatao Guo, observed that without an NLS, the LtrA protein expressed from a construct containing a similarly codon-optimized LtrA ORF under the CMV promoter was localized to the nucleus shortly after transfection, likely before enough protein is made to aggregate, and most hLtrA was cytoplasmic 48 h after transfection. In addition, in order to observe nuclear localization, it was necessary to transfect only minimal amount of DNA (personal communication). A possible explanation is that LtrA is highly basic, and the positive charges on the hLtrA protein surface may have functioned as nuclear localization signals until the hLtrA protein starts to aggregate in the cytoplasm, preventing the protein from entering the nucleus. An authentic NLS (as the one in phLtrA) is likely more efficient for

directing the hLtrA protein into the nucleus, so that some protein may aggregate in the nucleus rather than in the cytoplasm where it is made.

Another indication of aggregation is that a low percentage (<1%) of cells transiently transfected with phLtrA had cytoplasmic localization of LtrA, usually with immunofluorescence signal several fold stronger than that of nucleus-localized LtrA (not shown), suggesting that these cells received more copies of phLtrA and had a higher level of hLtrA expression. At the same time, stable cell lines expressing hLtrA from integrated copies of phLtrA had substantially lower expression of the hLtA protein (in both immunofluorescence and Western analysis) and nuclear localization in all observed cells.

4.2 EXPRESSION OF THE INTRON RNA

To express the intron RNA, I first cloned the intron with flanking exon sequences into vector pEGFP-N1 (Clontech), under a CMV promoter, to generate pEGFP-N1-ΔORF (not shown). Using RT-PCR on total RNAs of transfected HEK 293 cells, I could detect unspliced precursor but not ligated exons (not shown). In order to find optimal conditions for intron expression, I designed a series of expression vectors using various promoters, including the human RNA polymerase I promoter (Pol I), the CMV promoter in conjunction with a nuclear retention sequence from U3 snRNA (Speckmann et al., 1999) fused to the 3'-end of E2 (Pol II), the U3 snRNA promoter (Pol II) with different 3' ends, and the U6 snRNA promoter (Pol III).

Figure 4.6 shows the different constructs used to express the Ll.LtrB ΔORF intron. Each construct contains the Ll.LtrB-ΔORF intron and flanking LtrB exon sequences (172-180 bp of E1 and 90 bp of E2). Construct pHHWT uses the human Pol I promoter. pcDNAΔORF/NL uses the CMV promoter and the U3 nuclear retention sequence. In pTZU3+27ΔORF, intron expression is driven by the U6 snRNA promoter. The two pBSU3ΔORF constructs both use the U3 snRNA promoter and are identical

except that the sequences downstream of E2 are different. pBSU3ΔORF/NL has the U3 nuclear retention sequence and a pA signal, and pBSU3ΔORF.500 has instead 500 bp of genomic sequence that immediately follows the human telomerase RNA (hTR) and has been shown to improve the telomerase RNA expression level even when the RNA was under the U3 snRNA promoter instead of its endogenous promoter (Fu & Collins, 2003).

To detect splicing, each intron expressing-construct was either cotransfected with phLtrA into HEK 293 cells or transfected into phLtrA stable cell lines. Total RNA was isolated 48 h post transfection, and splicing of Ll.LtrB RNA was assessed by RT-PCR. As illustrated in figure 4.7(a), right panel, primers used for both RT and PCR anneal to the two ends of the exons. Ligated exons resulting from RNA splicing give a smaller product than that from unspliced precursors. The left panel in figure 4.7(a) shows the representative RT-PCR results. pHHWT (Pol I), pcDNAΔORF/NL (Pol II), and pTZU3+27ΔORF (Pol III) were transfected into phLtrA stable cell lines #25 and #38, respectively. pHHWT was transfected into HEK 293 cells as a control. Precursors were readily detectable in all the constructs. However, ligated exons were only observed in stable hLtrA cell lines expressing the Pol I construct (pHHWT). The correct exon junction was confirmed by sequencing the RT-PCR products (not shown). In some other experiments, weak ligated exon bands were seen, but not consistently, in cells transfected with Pol II and Pol III constructs. However, those bands were present in too low amounts to be sequenced. Figure 4.7b shows a different experiment where in addition to these above three constructs, the two pBSU3 constructs were also tested. Neither gave ligated exons. For the pHHWT (Pol I) construct, cotransfection with phLtrA and transfection into phLtrA stable cell lines resulted in similar intron expression levels and splicing efficiency.

In addition, the hLtrA expression cassette was cloned into the pHHWT construct to make pHHWT-hLtrA (not shown). When transfected into HEK 293 cells, pHHWT-hLtrA gave similar intron expression and splicing efficiency as cotransfection of pHHWT and phLtrA in RT-PCR (not shown).

4.3 OPTIMIZING SPLICING—THE Mg^{2+} EFFECT

A key condition for optimum splicing could be free Mg^{2+} concentrations *in vivo*. The Ll.LtrB intron has Mg^{2+} optima of ~5 mM for protein-dependent splicing and 10-20 mM for reverse splicing into DNA *in vitro* (Matsuura et al., 1997; Saldanha et al., 1999). *In vivo*, Mg^{2+} concentration is also critical for group II intron splicing. Studies in the yeast mt system have shown that mutations in the putative membrane transport proteins MRS2 and Lpe10, which decrease intramitochondrial Mg^{2+}, strongly inhibited the splicing of all four mt group II introns, while having little effect on the splicing of group I introns or other mt RNA processing reactions (Gregan et al. 2001a, b). Although human cells contain a total of 30 mM Mg^{2+}, only a small fraction of that (1-2 mM) is free Mg^{2+} (Ausubel et al., 1994; Romani & Scarpa, 2000). Elevated free Mg^{2+} concentrations may promote splicing *in vivo*. Here I tried to increase Mg^{2+} concentration in growth medium to see whether splicing could be improved.

First, I tried adding 20 mM of extra Mg^{2+} (10 mM each of $MgCl_2$ and $MgSO_4$) to the standard growth medium, DMEM, used in my experiments, which already contains 0.82 mM Mg^{2+}. I also used the Leibovitz L-15 medium (ATCC, Manassas, VA), which contains higher Mg^{2+} (1.82 mM). Neither gave increased efficiency of LtrA-promoted splicing in RT-PCR (not shown).

Bojanowski & Ingber (1998) applied up to 65 mM Mg^{2+} in the culture medium to decondense the chromosomes in living cells. Upon addition of extra Mg^{2+} into the medium, the authors observed instant but reversible unfolding of chromosomes. Thus, in

addition to potentially facilitating splicing, high Mg^{2+} concentrations might help make genomic targets more accessible. However, these authors only exposed the cells to high Mg^{2+} for up to 90 sec and found all effects reversible.

I first tested how long cells could sustain high Mg^{2+} concentrations in the medium. I incubated HEK293 cells in growth medium supplemented with 80 mM of $MgCl_2$ for different lengths of time. Immediately after adding $MgCl_2$, the cells looked different morphologically-- visible nuclei and nucleoli disappeared, and the contour of cells became smoother. After about an hour, both nuclei and nucleoli started to reappear, and cells assumed a more (but not completely) normal morphology. When high Mg^{2+}-containing medium was replaced by fresh medium without extra Mg^{2+}, the nuclei and nucleoli reappeared within seconds. Notably, after being exposed to 80 mM of Mg^{2+} for 4 h, the cells recovered and divided normally when placed overnight in the regular medium, demonstrated by lack of increased cell death or detectable morphological changes compared to untreated cells. These observations suggest that increased Mg^{2+} concentration in medium can result in at least a transient increase in intracellular Mg^{2+} concentration.

Unexpectedly, when I incubated 293 cells transfected with the Pol I intron construct with 80 mM Mg^{2+} for 1 h and 2 h, the intron spliced independently of the LtrA protein (Figure 4.8(a), lanes 1-4). The Mg^{2+}-induced splicing seems to be more efficient than that in LtrA-expressing cells at low Mg^{2+} (Figure 4.8(a), lanes 3, 4, and 5). Mg^{2+} increased splicing efficiency in LtrA-expressing cells to a similar level (lanes 3, 4, 7, and 8). RT-PCR of the same RNA samples with hLtrA primers proved that there was no LtrA contamination in Mg^{2+}-treated cells (Figure 4.8(b)). In order to exclude the possibility of intron splicing *in vitro* during preparation in the presence of 80 mM of Mg^{2+}, I washed

the cells with 0.5 M EDTA and included 80 mM EDTA throughout the preparation precedure.

The Mg^{2+}-induced effect could result from either self-splicing of the intron or intron splicing requiring some other cellular proteins. Potentially, in cells with genes containing inserted group II introns, conditional disruptions might be obtained by simply increasing the ambient Mg^{2+} concentration instead of providing the LtrA protein.

4.4 PRILIMINARY EVIDENCE THAT LTRA EXPRESSED IN MAMMALIAN CELLS HAS RT ACTIVITY

In addition to splicing activity, I wanted to know whether the hLtrA expressed in mammalian cells has RT activity. It is possible that the bacterial protein might be post-translationally modified in mammalian cells in ways that decrease its activities. Nuclear lysates were prepared by breaking the cell membrane with hypotonic buffer and homogenization (Eperon & Krainer, 1993; see section 6.2). RT assays were carried out both with the artificial template-primer substrate poly(rA)/oligo$(dT)_{18}$ and with a natural substrate mimic denoted Ll.LtrB/E2+10 (Wank et al., 1999; San Filippo & Lambowitz, 2002; Cui et al., 2004). The latter consists of the Ll.LtrB RNA (*in vitro* transcript containing the Ll.LtrB-ΔORF intron and flanking exons) with an annealed 20-mer DNA primer (E2+10), whose 3' end corresponds to that of the cleaved bottom strand normally used as the primer for TPRT.

Figure 4.9 shows RT activities in nuclear lysates from different cells. The assays were performed twice on the same batch of nuclear lysate with reproducible data. The nuclear lysates were made at approximately the same cells/volume ratio to help normalize the total protein level. In the Ll.LtrB/E2+10 RT assay (Figure 4.9(a), left panel), cells expressing LtrA had substantial RT activity compared to untransfected 293 cells under high salt conditions (450 NMT: 450 mM NaCl, 5 mM $MgCl_2$, and 40 mM Tris-HCl; pH

7.5; chapter 2). In 100 NMT (100 mM NaCl, 5 mM MgCl2, and 40 mM Tris-HCl, pH 7.5), all cells had similar activity, likely from endogenous polymerases. Although transient expression of hLtrA generates much more protein than either of the two stable cell lines (Figure 4.4), samples "hLtrA#25" and "hLtrA" have similar activity, possibly because a large portion of the hLtrA protein in transiently transfected cells aggregates, as suggested by the previous immunofluorescence and Western assays.

The poly(rA)/$(dT)_{18}$ assay was carried out in 100 NMT conditions. Cells expressing LtrA showed only background level of activity in the polyrA/$(dT)_{18}$ assay, similar to untransfected 293 cells. In addition, when purified LtrA was mixed with HEK 293 lysate, the poly(rA)/$(dT)_{18}$ activity was decreased to 20% of that without the cell lysate, indicating that the cell lysate is inhibitory to the reaction.

These experiments only qualitatively demonstrate that there was RT activity associated with hLtrA and do not indicate the specific activity. Whether or not the protein has full RT activity is still an open question.

4.5 PLASMID TARGETING AND HOMOLOGOUS RECOMBINATION WITH THE *LACZ* REPEAT CONSTRUCT

For group II intron targeting, the intron splices and forms RNPs with LtrA, recognizes, and inserts into the top strand of a DNA target, whereas LtrA cleaves the bottom strand and initiates reverse transcription. Finally, cellular mechanisms complete the reaction by second-strand synthesis and gap repair. The results above establish that the splicing step can occur in mammalian cells. Although it would be ideal if we could detect the end product (complete insertion) of targeting, it is desirable to detect cleavages, indicating that RNPs are functional *in vivo*. And cleavage, as an intermediate step, would occur more frequently than full reverse splicing (Dickson et al., 2004).

Here, I describe a plasmid system using *lacZ* as a reporter to detect potential cleavages by Ll.LtrB RNPs expressed in human cells. Double-strand breaks, such as those generated by intron RNPs on target sites, stimulate homologous recombination in mammalian cells. Although at the genome level, homologous recombination is a low-efficiency event, it is very efficient on plasmid and can occur with homologies as little as 14 bp (Rubnitz & Subramani, 1984). I designed a system similar to that used in Epinat et al. (2003), where the authors demonstrated that with a direct repeat configuration (Figure 4.10(b)), *lacZ* function could be restored when a double-strand break was generated by a homing endonuclease that site-specifically cleaves between the repeats.

A reporter construct was made so that the middle 300 bp of the 3-kb *lacZ* gene repeats itself with a wild-type Ll.LtrB target sequence (45 bp) and some flanking sequence in between (Figure 4.10(a), (b)). After cleavage occurs, cellular exonucleases digest the free ends from 5' to 3' end (resection). The repeat sequences will anneal to each other after becoming single-stranded, and extra sequences will be degraded and gaps and nicks will be fixed to form a complete copy of the *lacZ* gene.

For a negative control, I transfected HEK 293 cells with the reporter plasmid alone. As positive controls, the reporter plasmid was either linearized by using *Bsr*GI sites between the repeats, or incubated with RNPs under conditions that give cleavage and reverse transcription (with or without dNTPs, both gave similar results). After phenol extraction, both plasmids were transfected into the cells. Fourty-eight hours after transfection, cells were fixed and stained for β-galactosidase activity. Some cells (about 0.1%) in the negative control turned blue, indicating low background level of homologous recombination (the tranfected DNA was prepared using a CsCl gradient and should mostly be supercoiled, free of nicks or breaks; Figure 4.10(c), left). It is worth noting that given the very low frequency of the spontaneous homologous recombination

events, the chance for more than one plasmid in each cell to go through the process is very small. So it is reasonable to assume that the blue staining resulted from a single copy of *lacZ*. The linearized plasmid gave about 30-40% blue cells, approaching the transfection efficiency, implying at least one of the copies of linearized DNA that entered each cell had undergone homologous recombination. The RNP-treated sample gave about 2% blue cells, at least 10 times less efficient than the linearized sample, indicating that the RNPs do not cleave 100% of the target sites *in vitro*.

However, I tried to co-transfect intron and hLtrA expression vectors into the cells with the reporter plasmid, I did not observe increased homologous recombination as a result of RNPs cleaving the reporter. Increasing RNP production and activities by optimizing both the protein and intron RNA expression is crucial.

4.6 DISCUSSION

In this chapter, I showed that the Ll.LtrB intron and its IEP could be expressed in human cells for RNP production. I found that the expressed IEP has both maturase (RNA splicing) and RT activities.

To express the LtrA protein in mammalian cells, it was necessary to resynthesize the ORF using human-preferred codons. Human cells tolerate at least moderate levels of constitutively expressed hLtrA protein so that stable cell lines expressing hLtrA were selected (Figure 4.4). hLtrA with a C-terminal NLS fusion expressed both transiently and stably, is localized mainly in the nucleus (Figure 4.5). Splicing of the Ll.LtrB intron RNA was detected at least with the human RNA polymerase I promoter-driven constructs (Figures 4.7). In addition to splicing, hLtrA also has RT activity (Figure 4.8), which is essential for targeting.

However, several lines of evidence suggest that hLtrA may aggregate in the cells. First, I found a band several times larger than LtrA in Western analysis in transiently

transfected cells (with higher hLtrA expression level) disappears when more SDS-containing buffer was used to dissolve the cells. This band appeared in variable amounts relative to hLtrA in different experiments. Second, some cells transiently transfected with phLtrA showed cytoplasmic localization when stonger immunostaining signal was observed, indicating of increased hLtrA expression. And hLtrA was localized to the nucleus in COS-7 cells, where phLtrA can replicate, but avoided the dense nucleoli. Lastly, although transient transfection resulted in higher hLtrA expression than stable hLtrA-expressing cell lines, similar RT activity was detected in both transiently and stably transfected cells. Taken together, these findings suggest that hLtrA is likely to aggregate in human cells when expression is relatively high, as in transient transfections. The expression system needs to be optimized to express the maximum amount of soluble protein. It would also be worthwhile to assess whether post-translational modifications negatively affect LtrA function. Both splicing and RT activities could be improved by preventing deleterious modifications.

The Ll.LtrB RNA was expressed from pol I, pol II, and pol III promoters. However, only Ll.LtrB RNA expressed from a pol I promoter showed reporducible LtrA-dependent splicing *in vivo*. For all constructs tested, the level of expressed Ll.LtrB RNA was so low that the RNA beingwas only detectable by RT-PCR, but not by Northern hybridizations. The similar intron RNA expression levels from different promoters, despite their different strengths for transcription, raises the possibility that some cellular mechanism, such as microRNAs, RNAi, or nucleases, may control how much the Ll.LtrB RNA can be maintained in the cell. This aspect should be investigated along with other ways to optimize intron production.

By testing how extracellular Mg^{2+} can affect intron splicing, I found surprisingly that the Ll.LtrB intron expressed in HEK 293 cells spliced without LtrA when high

concentrations of Mg^{2+} were added to the growth medium (Figure 4.8(a)). An attractive aspect of this Mg^{2+}–induced splicing is that it might be used to obtain conditional disruptions for genes containing group II introns. Instead of introducing LtrA into the cells, merely adding Mg^{2+} into the medium will induce splicing of the inserted intron.

Finally, I established a plasmid reporter system to detect RNP cleavages of a target site. Essentially, once RNPs cleave the target site between repeated sequences of the reporter and produce double-strand breaks, 5' resection occurs from the break points and repeated sequences become two complementary single strands. Homologous recombination through single-strand annealing will generate a functional copy of the reporter. Here, I used *lacZ* as the reporter gene and proved the system worked with both negative and positive controls. However, I have not observed homologous recombination stimulated by breaks resulting from *in vivo* expressed Ll.LtrB RNPs cleaving the target site. This system can also be used to measure RNP activity by incubating RNPs with reporter plasmid *in vitro* before transfecting the plasmid into the cells to allow homologous recombination.

4.7 SUMMARY

Both the LtrA protein and the Ll.LtrB intron RNA were expressed in human cells. The LtrA protein is mostly localized in the nucleus and has both maturase and RT activities. The protein may be aggregated under some conditions. Notably, I also observed LtrA-independent splicing of the Ll.LtrB intron *in vivo*, when 80 mM $MgCl_2$ was added to the growth medium in human cells. In addition, a plasmid assay was developed to detect RNP cleavage activity *in vivo*.

Figure 4.1 Constructs for expressing LtrA in human cells.

(a) pCMV-LtrA contains the LtrA ORF under a CMV promoter and with C-terminal NLS fusion. pA is a poly A signal, and ori is an *E. coli* replication origin. (b) pCMV-IVS is an intermediate construct that contains a short (132 bp) spliceosomal intron (IVS) cloned downstream of the CMV promoter to improve protein expression (Le Hir et al., 2003). (c) pCMV-IVS-LtrA has LtrA cloned downstream of the IVS and also has the C-terminal NLS. (d) pKS-hLtrA is an intermediate construct that contains the full-length codon-optimized LtrA sequence. There is no promoter upstream of hLtrA ORF to drive transcription. (e) phLtrA is the expression vector for hLtrA with the ORF cloned downstream of the CMV promoter and the IVS sequence and an SV40 NLS fused to the C-terminus of the protein. (f) pLtrA was constructed by swapping the hLtrA ORF with PCR-amplified LtrA ORF. (g) phLtrAΔNLS is identical to phLtrA except that the NLS was deleted.

Figure 4.1

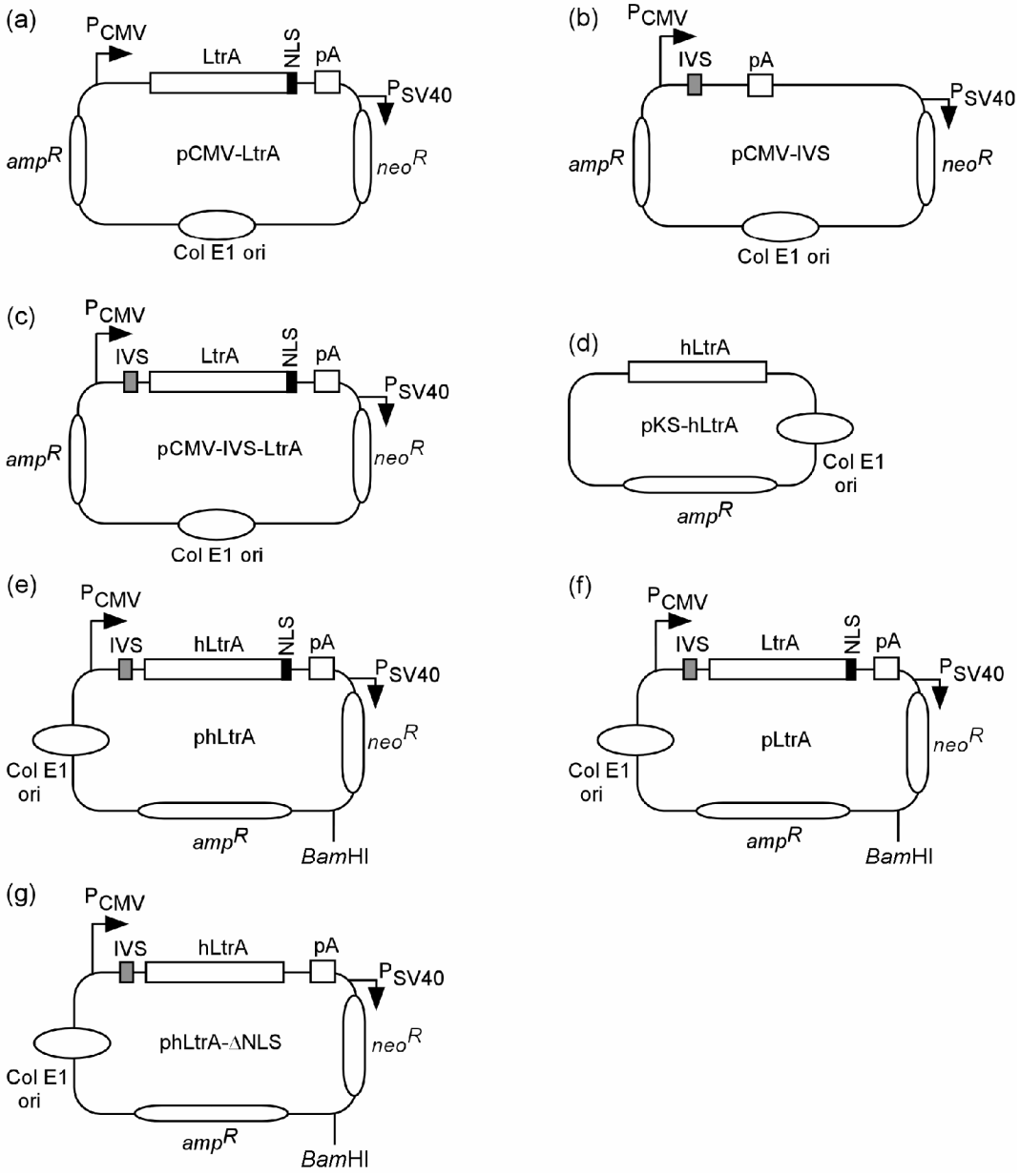

(a)
P CMV
LtrA
NLS
pA
P SV40
amp R
pCMV-LtrA
neo R
Col E1 ori
(b)
P CMV
IVS
pA
P SV40
amp R
pCMV-IVS
neo R
Col E1 ori
(c)
P CMV
IVS
LtrA
NLS
pA
P SV40
amp R
pCMV-IVS-LtrA
neo R
Col E1 ori
(d)
hLtrA
pKS-hLtrA
Col E1 ori
amp R
(e)
P CMV
IVS
hLtrA
NLS
pA
P SV40
phLtrA
neo R
Col E1 ori
amp R
BamHI
(f)
P CMV
IVS
LtrA
NLS
pA
P SV40
pLtrA
neo R
Col E1 ori
amp R
BamHI
(g)
P CMV
IVS
hLtrA
pA
P SV40
phLtrA-ΔNLS
neo R
Col E1 ori
amp R
BamHI

Figure 4.2 DNA sequence of hLtrA ORF.

The Kozak sequence is in bold, italic, and lower case at the N-terminus. The translational initiation codon ATG is in bold immediately following the Kozak sequence. At the C-terminus, the SV40 NLS is in bold and italic lower case immediately preceding the stop codon TAA. Numbering is on the right of the sequence.

Figure 4.2

gaatt***ccacc*ATG**AAGCCCACCATGGCCATCCTGGAGCGCATCAGCAAGAACTCACAGGAGAACATCGACGAGGT 75

GTTCACCCGCCTGTACCGCTACCTGCTGCGCCCCGACATCTACTACGTGGCCTACCAGAACCTGTACAGCAACAA 150

GGGCGCCAGCACCAAGGGCATCCTGGACGACACCGCCGACGGCTTCAGCGAGGAGAAGATCAAGAAGATCATCCA 225

GAGCCTGAAGGACGGCACCTACTACCCCCAGCCCGTGCGCCGCATGTACATCGCCAAGAAGAACAGCAAGAAGAT 300

GCGCCCCCTGGGCATCCCCACCTTCACCGACAAGCTGATCCAGGAGGCCGTGAGGATCATCCTGGAGTCCATCTA 375

CGAGCCCGTGTTCGAGGACGTGAGCCACGGCTTCCGCCCCCAGCGCAGCTGCCACACCGCCCTGAAGACCATCAA 450

GCGCGAGTTCGGCGGCGCCAGGTGGTTCGTGGAGGGCGACATCAAGGGCTGCTTCGACAACATCGACCACGTGAC 525

CCTGATCGGCCTGATCAACCTGAAGATCAAGGACATGAAGATGAGCCAGCTGATCTACAAGTTCCTGAAGGCCGG 600

CTACCTGGAGAACTGGCAGTACCACAAGACCTACAGCGGCACCCCCCAGGGCGGCATCCTGAGCCCCCTGCTGGC 675

CAACATCTACCTGCACGAGTTGGACAAGTTCGTGCTGCAGCTGAAGATGAAGTTCGACCGCGAGAGCCCCGAGCG 750

CATCACCCCCGAGTACCGCGAGCTGCACAACGAGATCAAGCGCATCAGCCACCGCCTGAAGAAGCTGGAGGGCGA 825

GGAGAAGGCCAAGGTCCTGCTGGAGTACCAGGAGAAGCGCAAGAGGCTGCCCACCCTCCCCTGCACCAGCCAGAC 900

CAACAAGGTGCTGAAGTACGTGCGCTACGCCGACGACTTCATCATCAGCGTGAAGGGCAGCAAGGAGGACTGCCA 975

GTGGATCAAGGAGCAGCTGAAGCTGTTCATCCACAACAAGCTGAAGATGGAGCTGAGCGAGGAGAAGACCCTGAT 1050

CACCCACAGCAGCCAGCCCGCCCGCTTCCTGGGCTACGACATCCGCGTGCGCCGCAGCGGCACCATCAAGCGCAG 1125

CGGCAAGGTGAAGAAGCGCACCCTGAACGGCAGCGTGGAGCTGCTGATCCCCCTGCAGGACAAGATCCGCCAGTT 1200

CATCTTCGACAAGAAGATCGCCATCCAGAAGAAGGACAGCAGCTGGTTCCCCGTGCACCGCAAGTACCTGATCCG 1275

CAGCACCGACCTGGAGATCATCACCATCTACAACAGCGAGCTGAGGGGCATCTGCAACTACTACGGCCTGGCCAG 1350

CAACTTCAACCAGCTGAACTACTTCGCCTACCTGATGGAGTACAGCTGCCTGAAGACCATCGCCTCCAAGCACAA 1425

GGGCACCCTGAGCAAGACCATCTCCATGTTCAAGGACGGCAGCGGCAGCTGGGGCATCCCCTACGAGATCAAGCA 1500

GGGCAAGCAGCGCCGCTACTTCGCCAACTTCAGCGAGTGCAAGTCCCCCTACCAGTTCACCGACGAGATCAGCCA 1575

GGCCCCCGTGCTGTACGGCTACGCCCGCAACACCCTGGAGAACCGCCTGAAGGCCAAGTGCTGCGAGCTGTGCGG 1650

CACCAGCGACGAGAACACCAGCTACGAGATCCACCACGTGAACAAGGTGAAGAACCTGAAGGGCAAGGAGAAGTG 1725

GGAGATGGCCATGATCGCCAAGCAGCGCAAGACCCTGGTGGTGTGCTTCCACTGCCACCGCCACGTGATCCACAA 1800

GCACAAG***gtcgacccccccaagaagaagaagcgcaaggtggcc*TAA**

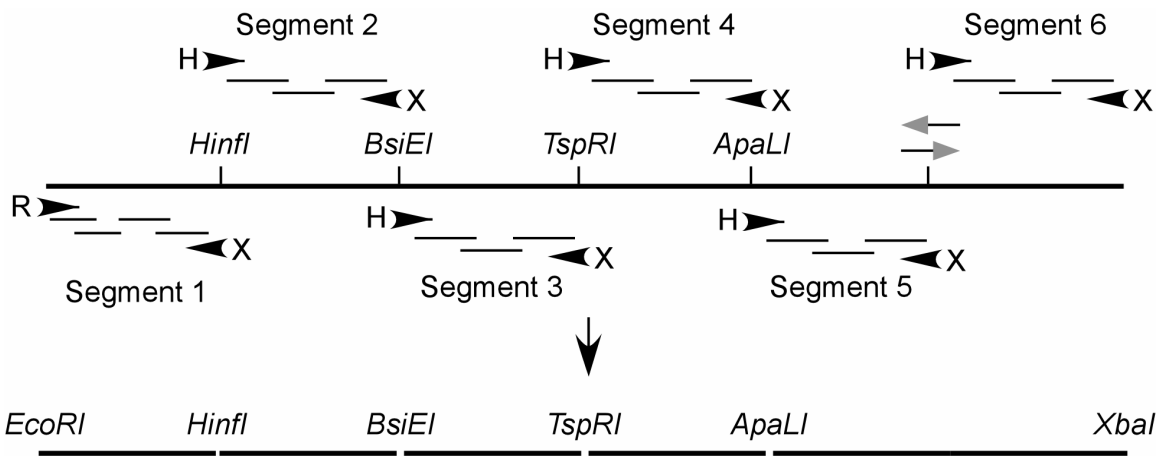


Figure 4.3 Schematic of cloning strategies for constructing the hLtrA ORF.

Horizontal bars represent overlapping synthetic DNA oligonucleotides used as templates in PCR reactions. Arrowheads showed smaller primers used to amplify each segment. Letters next to arrowheads represent restriction sites that were later used in cloning. The letter at the end of each short arrow R: *Eco*RI; H: *Hind*III; X: *Xba*I. Grey arrows represent the two complementary oligonucleotides used to amplify segments 5 and 6 into one piece. Each segment was flanked by two restriction sites that are unique within the fragment and were later used to ligate all the pieces together.

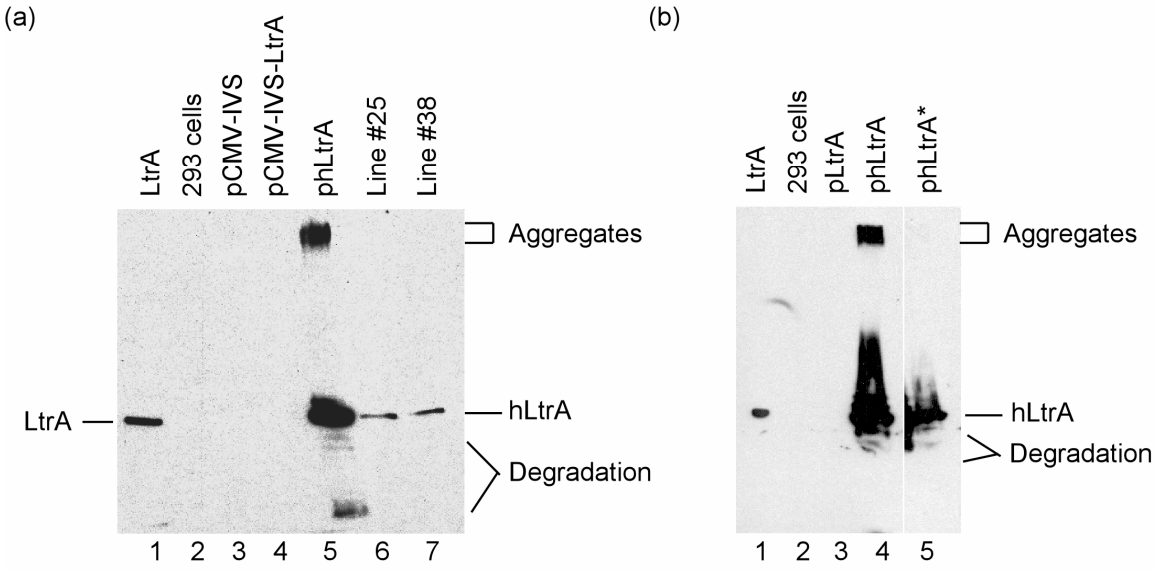


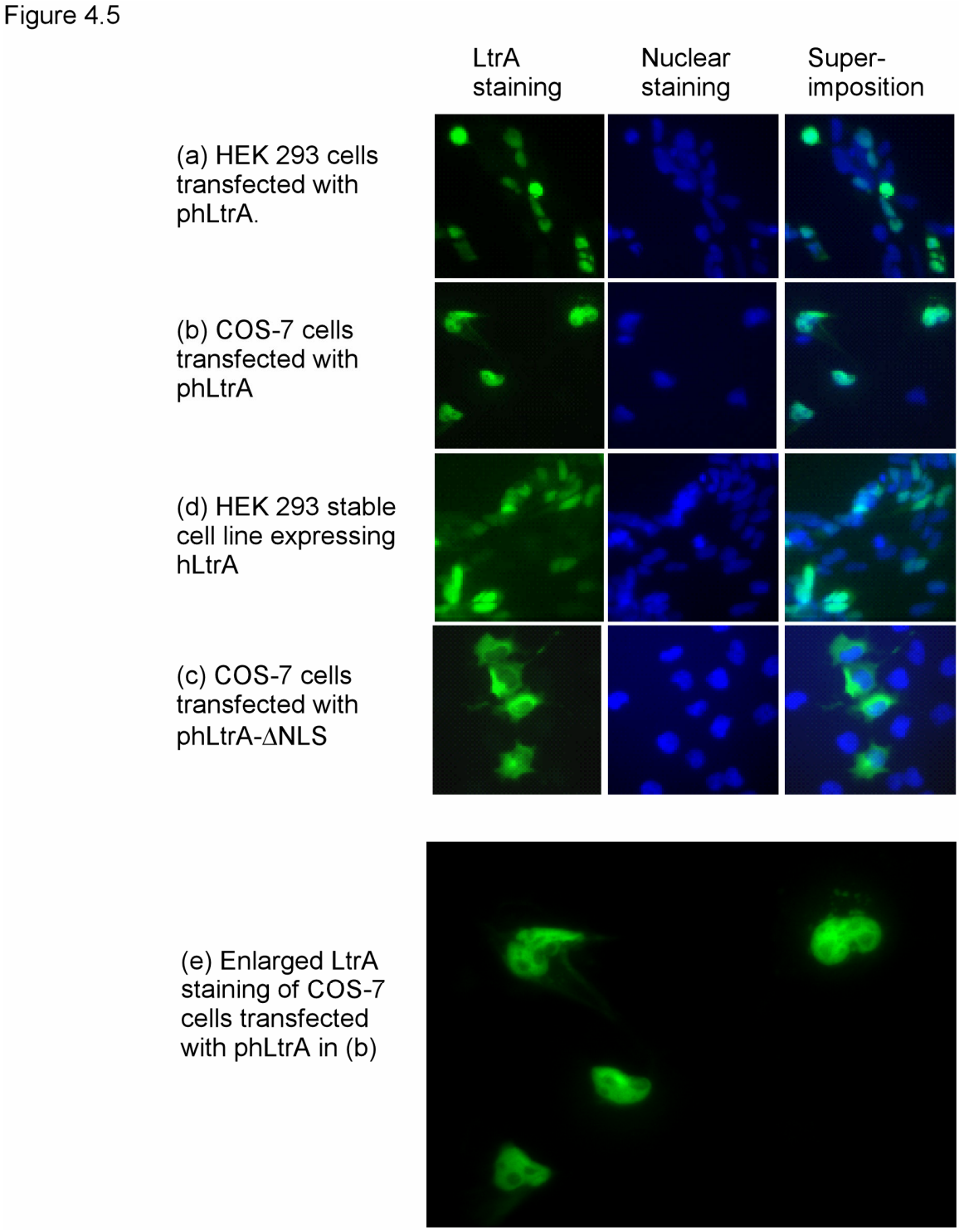
Figure 4.4 Western analysis showing expression of the codon-optimized LtrA construct.

(a) hLtrA expresses well in transiently and stably transfected HEK 293 cells (lanes 5 to 7), whereas no hLtrA expression was detected in untransfected 293 cells (lane 2), 293 cells transfected with pCMV-IVS (lane 3), or pLtrA (LtrA without codon optimization; lane 4)**.** LtrA protein purified from *E. coli* was run as a positive control and size marker (lane 1). (b) hLtrA expression in transiently transfected 293 cells compared to no detectable expression from pLtrA, which is identical to phLtrA except that the ORF is not codon-optimized (lane 3) or untransfected 293 cells (lane 2). Lanes 4 and 5 were from the same sample. Lane 4 was loaded with the supernatant of the sample after boiling in SDS-containing loading buffer, and lane 5 was loaded with the pellet portion after boiling again with more SDS buffer. Lane 1 is LtrA purified from *E. coli*.

Figure 4.5 Localization of hLtrA.

An anti-LtrA antibody was used in immunofluorence assay, where the secondary antibody was conjugated with FITC. (a) to (d) show the localization of hLtrA in cells and from constructs indicated in the figure. The left panel is LtrA immuno-staining; the middle panel is nuclei stained with Hochest dye; and the right panel is the superimposition of the first two. (e) is an enlarged image of the left panel of (b).

Figure 4.5


LtrA staining
Nuclear staining
Super-imposition
(a) HEK 293 cells transfected with phLtrA.
(b) COS-7 cells transfected with phLtrA
(d) HEK 293 stable cell line expressing hLtrA
(c) COS-7 cells transfected with phLtrA-ΔNLS
(e) Enlarged LtrA staining of COS-7 cells transfected with phLtrA in (b)

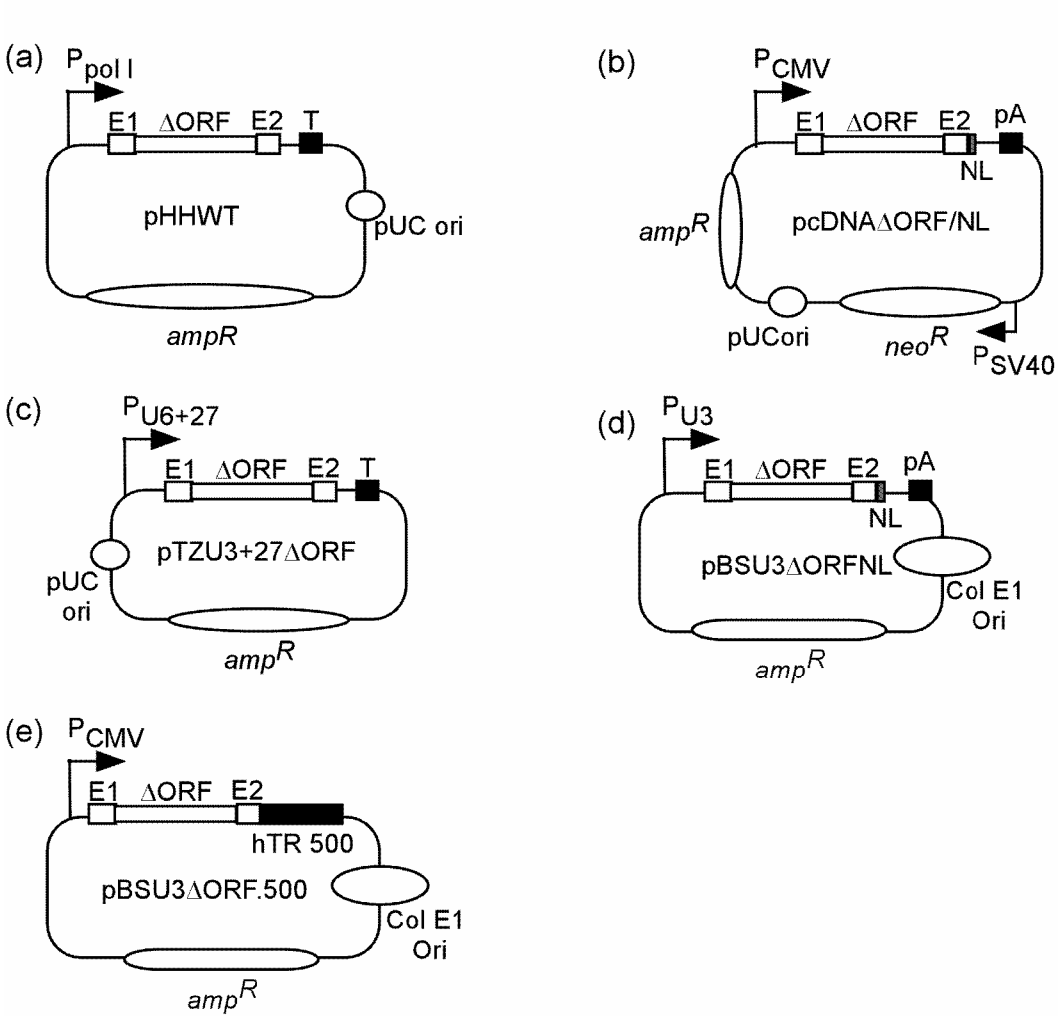


Figure 4.6 Constructs used to express the intron RNA.

E1 and E2 denote exons 1 and 2, respectively. ΔORF represents the ΔORF version of the Ll.LtrB intron. pA is a poly A signal. T is a transcriptional terminator. NL is the nuclear localization sequence from the U3 snRNA. An arrow shows the position of the indicated promoter. hTR500 in (e) represents the 500 bp non-transcribed sequence downstream of the human telomerase RNA. Ori is a replication origin. Among these constructs, Ms. Junhua Zhao made pcDNAΔORF/NL and pTZU6+27ΔORF.

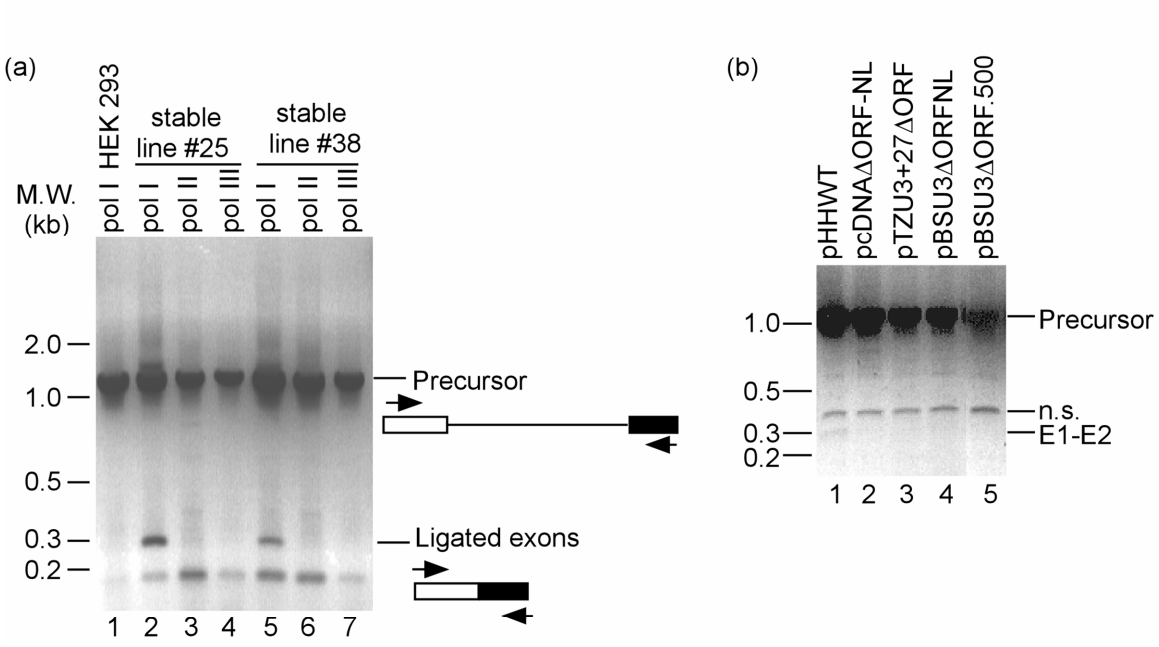


Figure 4.7 RT-PCR results of intron expressed from the constructs in figure 4.6.

(a) Splicing of different constructs in 293 cells or hLtrA stable cell lines. Pol I, Pol II, and Pol III refer to pHHWT, pcDNAΔORFNL, and pTZU6+27ΔORF, respectively. (b) *In vivo* splicing in HEK 293 cells with pBSU3ΔORFNL and pBSU3ΔORF.500. n.s. stands for nonspecific amplification.

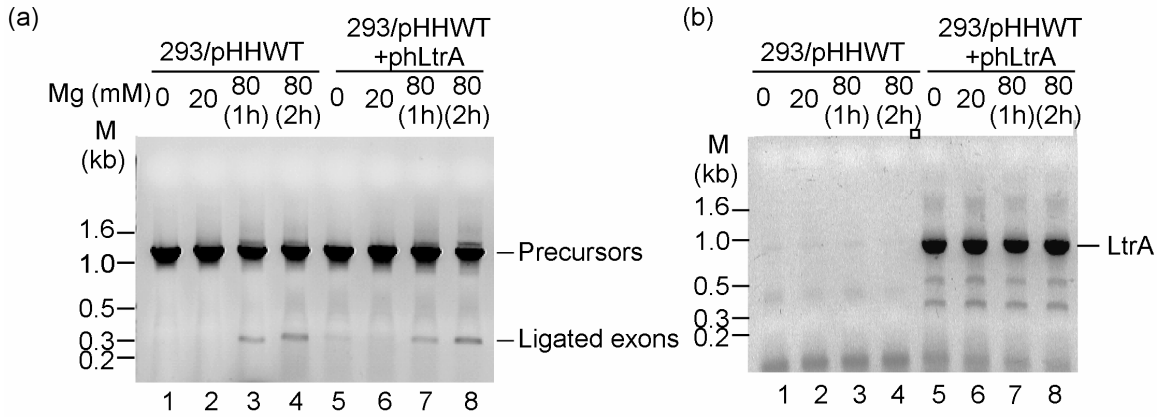


Figure 4.8 Mg^{2+}-induced splicing of the Ll.LtrB intron.

(a) RT-PCR demonstrating Mg^{2+}-induced splicing in human cells. Cells transfected with pHHWT (lanes 1 to 4) or cotransfected with pHHWT and phLtrA were incubated with different concentrations of Mg^{2+}, 0, 20, or 80 mM, respectively (lanes 5 to 8). After cells were incubated in 80 mM of Mg^{2+} for 1 h or 2 h, Mg^{2+} was diluted to 20 mM with fresh medium and cells were all harvested after 3 h of the initial addition of Mg^{2+}. (b) RT-PCR of all the RNA samples in the left panel with hLtrA primers.

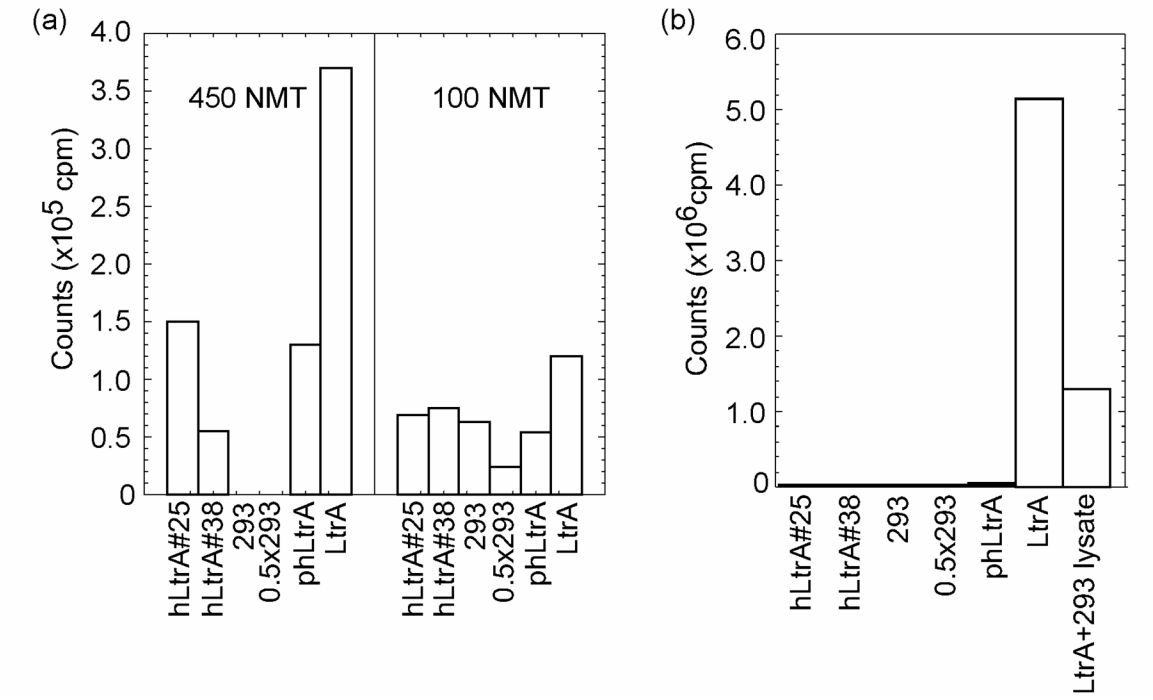


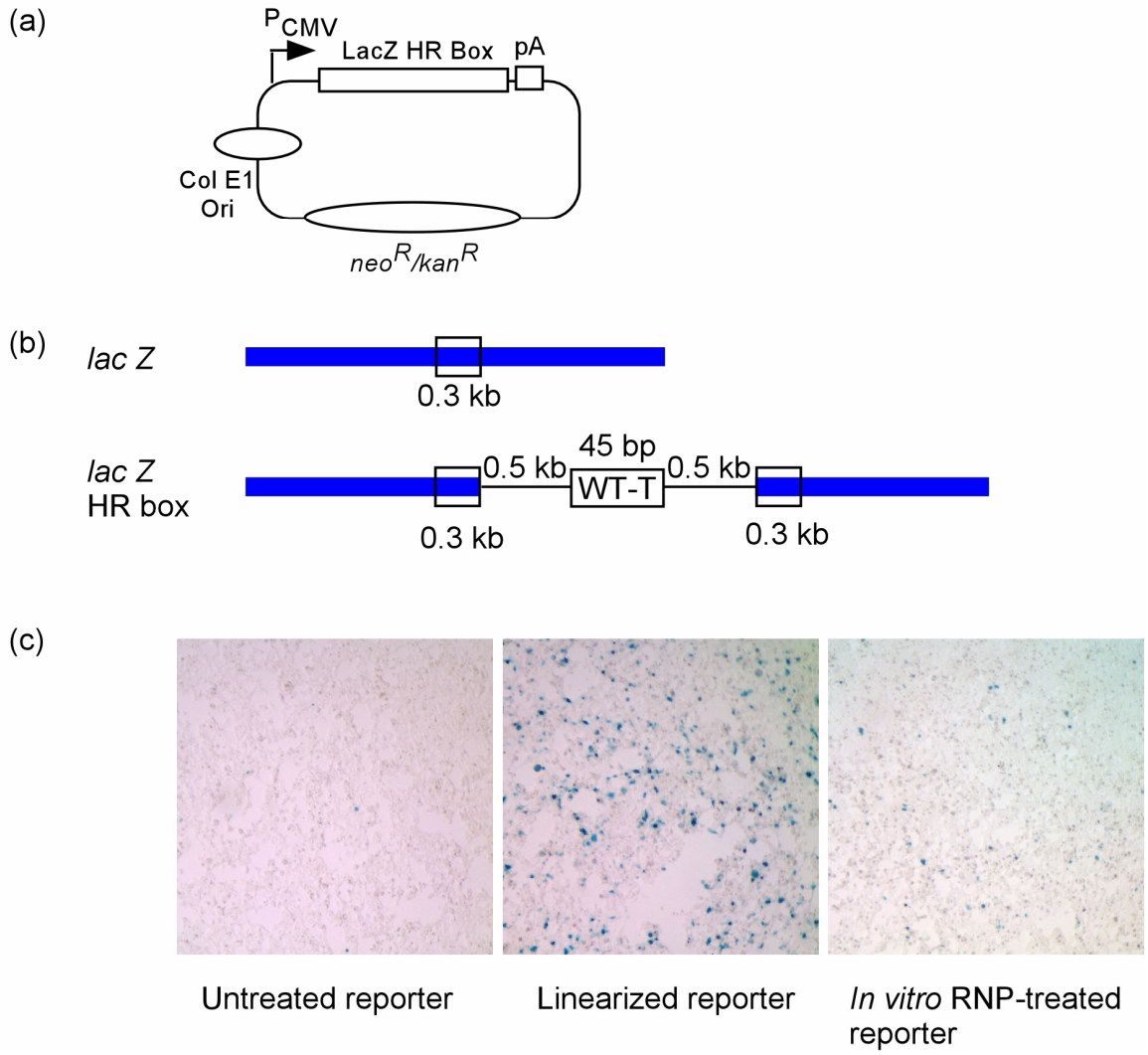
Figure 4.9 RT assays using nuclear lysates.

hLtrA#25 and hLtrA#38 are HEK 293 cell lines stably expressing hLtrA. 293 and 293-1/2 represent non-transfected cells, and in 293-1/2, only half amount of the lysate was used. LtrA stands for 0.5 μg of purified LtrA protein. LtrA+293 lysate is the mixture of purified LtrA and 293 lysate. (a) The Ll.LtrB/E2+10 RT assay. Reaction media were as indicated. 450 NMT stands for 450 mM NaCl, 5 mM $MgCl_2$, and 40 mM Tris-HCl (pH 7.5), and 100 NMT is identical except that the NaCl concentration is 100 mM. (b) Poly(rA)/oligo$(dT)_{18}$ assay (see section 6.1.10).

Figure 4.10 A plasmid system for detecting homologous recombination stimulated by RNP cleavage of a target site.

(a) The reporter plasmid. The *lacZ* HR box contains a CMV promoter, a *lacZ* ORF with the middle portion (from 1.3 kb to 1.6 kb downstream of the ATG codon) directly repeating itself, with the two repeats separated by an insertion including 45 bp of wild-type target site and 500 bp flanking sequences on either side. (b) A comparison of the wild-type *lacZ* gene and the modified *lacZ* gene in the HR box. The open boxes represent the 300-bp repeated sequence. (c) shows that the system works to detect homologous recombination. HEK 293 cells were transfected with untreated reporter plasmid, reporter plasmid linearized between the repeats by *Bsr*GI digestion, and reporter plasmid preincubated with RNP particles *in vitro*. After transfection, the cells were fixed and stained for β-galactosidase activity (6.2.13). Pictures were taken with 6x magnification.

Figure 4.10


(a)
P CMV
LacZ HR Box
pA
Col E1
Ori
neo R/kan R
(b)
lac Z
0.3 kb
lac Z
HR box
45 bp
0.5 kb
WT-T
0.5 kb
0.3 kb
0.3 kb
(c)
Untreated reporter
Linearized reporter
In vitro RNP-treated reporter

Chapter 5: Using purified RNP to target the mammalian genomes

An alternative way of using group II introns in gene targeting is to deliver purified RNPs directly to cells. In order to target the chromosomes, group II intron RNPs must be in proximity of their targets when the chromosomes are structurally accessible. Practically, it is feasible to obtain large quantities of highly active RNPs either from *E. coli* expression or by reconstituting them using purified LtrA protein and self-spliced intron RNA. Unlike plasmid DNA used for intorn expression, which could integrate randomly into the genome, RNPs turn over relatively fast and are less likely to have long-lasting adverse effects on the cells. However, disadvantages of using RNPs include low transfection efficiency and short shelf life, thus less exposure to the targets than in the case of expressing RNPs *in vivo*. Guo et al. (2000) showed that RNP particles could be delivered into human cells by lipisome-mediated transfection and were able to insert into a plasmid-born target site. However, the reaction was very inefficient and required nested PCR to detect the insertion product. In addition, it was not clear where in the cell the targeting event occurred. Obviously, to target chromosomes, RNPs must enter the nucleus and stay active there. A major obstacle of using RNPs for genomic disruption is how to introduce RNP particles into the nucleus efficiently. Another factor to be determined is whether RNPs could access targets within chromatin structures.

In this chapter, I describe *in vitro* experiments showing that the Ll.LtrB intron can target mammalian genomes and *in vivo* evidence that a modified Ll.LtrB intron inserted into an rDNA target site. I first showed that RNPs could insert into even single-copy target sites in purified genomic DNAs. Then I showed RNPs inserted successfully into target sites in nuclear lysates, where target genes are in chromatin structures. Additional buffer was not necessary in some reactions, indicating nuclear lysates contain sufficient

Mg^{2+} and proper ionic strength for targeting reactions to occur. I also tested different ways of introducing purified RNPs into cells, including injection and electroporation. Lastly, I detected an insertion event into the rDNA gene in HEK 293 cells with RNP electroporation.

5.1 *IN VITRO* GENOMIC TARGETING WITH RNPS

To demonstrate the feasibility of using RNPs to target the human genome, I tested whether group II intron RNPs could integrate into specific target sites in the context of a complex genome *in vitro*.

5.1.1 Targets with different copy numbers

I chose several chromosomal targets with different copy numbers to test the sensitivity of group II intron gene targeting. Introns against each target were either selected using a random library (Guo et al., 2000) or designed using a computer algorithm (Perutka et al., 2004).

The human LINE-1 (L1) retrotransposon is a good target because there are about 50,000 copies of L1 sequences per human cell, which comprise apporximately 17% of the human genome. The majority of L1s have 5' deletions (Brouha et al., 2003). An efficient intron (L1-2458a, 115% wild-type efficiency in an *E. coli* plasmid assay) was selected from the random intron library constructed by Guo et al. (2000). This intron targets the middle portion of the 6-kb L1 sequence (Figure 5.1; D'Souza & Lambowitz, unpublished).

rDNA genes have a medium copy number of 400 per diploid human cell (Yoon et al., 1995; Gecheva et al., 1996). The copy number is higher in multiploid cell lines. In addition to having multiple copies, rDNA genes have the advantage of being actively transcribed in fast growing cells and thus may be highly accessible to RNPs. Several

introns were designed to target different locations in the 45S rDNA and tested using the *E. coli* mobility assay (Guo et al., 2000). The most active one (rDNA-16280a, 70% wild-type efficiency) was used in the experiments below (Figure 5.1).

Two low copy number targets are mouse hypoxanthine phosphoribosyl transferase (mHPRT) and Herpes Simplex Virus (HSV) thymidine kinase (TK) gene, both of which are selectable (Mansour et al., 1988; Horie et al., 1994; Cannon et al., 1999). The HPRT gene is an endogenous gene located on the X-chromosome, and thus present in single-copy in male diploid cells, such as mouse ES cell lines. An efficient intron (mHPRT-144s, 75% wild-type efficiency) was designed using the computer algorithm to target the sense strand of the gene (Figure 5.1).

The HSV-TK gene was introduced into HEK 293 and NIH 3T3 cell lines on a Murine Leukemia Virus (MLV) vector by Bruce Sullenger's laboratory at Duke University, and single-copy clones were used in the experiments described in this chapter. I selected an intron (TK-61s, 105% wild-type efficiency) with wild-type frequency against the HSV-TK gene from the random library (Guo et al., 2000).

5.1.2 *In vitro* TPRT-PCR with purified chromosomal DNAs

First, I wanted to find out whether group II intron RNPs were able to identify their targets in the complex mammalian genomes. I purified chromosomal DNA from HEK 293 cells and a clone of NIH 3T3/TK cells (stable cell lines derived from a mouse cell line, NIH 3T3, containing a single copy of the HSV-TK gene). RNPs were either purified from *E. coli* by affinity chromatography (rDNA) or reconstituted *in vitro* (L1, TK, and mHPRT) (see methods). RNPs against rDNA and L1 were incubated with the HEK 293 DNA, whereas RNPs against TK and mHPRT were incubated with the NIH 3T3/TK DNA in the target DNA-primed reverse transcription assay (TPRT; see methods). Following TPRT, PCR was used to detect insertion junctions with an intron primer and a

target site primer in each reaction (Figure 5.2(a)). Nested PCR was performed when necessary. All PCR-amplified bands of expected size were sequenced to confirm the correct junctions.

Figure 5.2(b) shows that incubation of L1 and rDNA RNPs with the HEK 293 DNA resulted in readily detectable 3' junctions. The 3' junction of the L1 insertion site (lane 3) only required one round of PCR, while nested PCR was necessary to detect the 3' insertion junction of the rDNA target (lane 1). Incubation of the wild-type RNP with the chromosomal DNA resulted in nonspecific amplifications (lanes 2 and 4). The 5' junction was more difficult to detect, presumably because the intron RNA may insert partially into the target to form only the 3' junction, and/or even when the insertion is complete, reverse transcription of the RNA insertion may not be complete (see Figure 1.4). In addition, non-optimal PCR conditions may also be a factor. In lane 7, only a faint band of L1 5' insertion junction was amplified by one-step PCR, unlike that of the 3' junction (lane 3). On the other hand, the 5' junction was readily detectable in rDNA samples (lane 5). However, I could not exclude that it resulted from template switching during PCR. All the RNP preparations contain small amount of contaminating intron-donor plasmid DNA, which shares up to 12 bp of homology with the target site (Figure 5.2(d)). When the GC content is high, *e.g.*, 100% GC in IBS sequences of the rDNA target, nested PCR would pick up low amounts of template-switching products. In fact, I was able to obtain the same band as that in lane 5 by including both the donor plasmid and the 293 DNA as templates in nested PCR (not shown).

Figure 5.2(c) shows that RNPs were able to target low copy number targets in mammalian genomes. NIH 3T3 DNA was incubated with TK and mHPRT RNPs, respectively. PCR amplified TK and mHPRT 3' junctions with respective target primers (lanes 1 and 4). Primers of one target site did not amplify the insertion junction in the

DNA treated with the other RNP preparation (lanes 2 and 3), demonstrating specific insertion. The mHPRT RNP gave a specific 5'-insertion junction (lanes 7 and 8). However, TK targeting did not give a detectable 5' junction (lane 5), which could be due to non-optimal PCR conditions for this primer/template combination.

It is worth noting that, in another experiment, when I used less concentrated RNP preparations and a different preparation of DNA, I was able to amplify the 3' junction of targeted mHPRT only marginally, and TK was not detectably inserted (not shown). In yet another experiment, I was able to obtain good amplifications of the 5' junction of targeted L1 (dramatically better than Figure 5.2(b), lane 7) using one-step PCR. So RNP concentration and DNA purity may greatly affect the targeting efficiency *in vitro*.

Together, these results showed that RNPs are able to insert specifically into mammalian genomic DNA targets, including single-copy genes.

5.1.3 *In vitro* TPRT with nuclear lysates

Next, I used nuclear lysates as targets in the TPRT assay to find out whether RNPs could access the chromosomal targets within chromatin structures. I made nuclear lysates by first treating the cells with a hypotonic buffer (10 mM HEPES and 10 mM KCl, pH 8.0) and then homogenizing (Eperon & Krainer, 1993). I avoided sonication of the nuclear pellets so that chromosomal DNA stayed relatively intact. Instead, three cycles of freezing and thawing were applied to the nuclear pellets to break the nuclear membrane. Complete separation of the nucleus from the cytoplasm is not critical for my purposes.

In TPRT assay, I incubated intron RNPs and nuclear lysate with 2 mM dNTPs and with or without Mg^{2+}-containing RT buffer. Unlike targeting L1 in the purified chromosomal DNA, detection of the 3' junction of L1 in nuclear lysate required nested PCR (lanes 5 and 6). Lane 7 shows the insertion junction was produced specifically by

RNPs against L1 but not by those against rDNA. The difference between lanes 5 and 6 is that RT buffer (40 mM Tris-HCl, pH 7.5, 10 mM KCl, 10 mM $MgCl_2$, 5 mM DTT) was included in the TPRT reaction of lane 5 but not in that of lane 6. Similar results in these two lanes suggest that additional Mg^{2+} is not necessary for targeting to occur in the lysate. The 5' junction was not detected for L1 in the lysate. Unlike the situation for targeting purified chromosomal DNA, nuclear lysates probably contains nucleases, which could potentially digest the inserted intron RNA, causing the RT reaction to run off the template and lowering the detection of 5' junction.

Nested PCR did not amplify the 3' junction of targeted rDNA specifically (lane 1) and gave a nonspecific band in the reaction of L1 RNP-treated lysate with rDNA primers (lane 2). In another experiment, I was able to use a different batch of *in vivo* RNP to target rDNA in lysate (not shown). The 5' junction was amplified, but again, it may have resulted from template switching (lane 3). I did not test the two low copy targets in the nuclear lysate.

5.2 DELIVERY OF RNPS BY INJECTION

Injection is the most certain way to deliver RNP into the nucleus. In collaboration with the Transgenic Mice Facility on campus, we tried to inject RNPs into fertilized mouse eggs. Ms. Shanon Maika performed the injections for me. However, a limiting factor is that one could only feasibly inject a relatively small number (about 100) of eggs at a time, which may not be sufficient to find rare events. In addition, RNP preparations, especially those more concentrated than 1 μg/μl, tend to clog the injection needles, further lowering the number of eggs and the amount of RNPs per egg being successfully injected.

We have tried to inject mHPRT and rDNA RNPs, both purified from *E. coli*, and neither resulted in targeting events that were detectable by nested PCRs. On the other

hand, others in the lab injected RNPs to *Xenopus* oocytes, zebra fish and *Drosophila* embryos and observed targeting events in plasmid-born targets, with efficiency directly related to the RNP concentration (Watanabe, Vernon, & Lambowitz, unpublished). These other systems use needles with larger diameters, hence no clogging, and allow larger injection volumes than in the case of mouse egg injection. Injection remains an attractive means for introducing RNPs into the nucleus. Improvement could be made in RNP preparation to increase both concentration and solubility.

5.3 FEASIBILITY OF USING ELECTROPORATION TO DELIVER RNPS INTO THE CELLS

Lipid-mediated transfection reagents are commonly used for delivery of nucleic acids into cells. However, the presence of protein greatly decreases transfection efficiency. Peptide vectors are useful in delivering proteins (Morris et al., 2001). However, I could not detect binding between the peptide vectors (MPG and chariot) and RNPs. Electroporation induces transient openings on the cytoplasmic membrane, which allows molecules of all sizes and types to enter the cell. A drawback of electroporation is that cell survival rate is usually much lower than lipid-mediated transfection.

Many groups have used electroporation to successfully deliver nucleic acid/protein complex, including influenza viral (Li et al., 1995) and DNA transposon complex (Reznikoff et al., 2004).

5.3.1 Testing electroporation conditions for RNP

To test whether electroporation delivers RNP into cells, I used an RNP preparation with an N-terminal GFP fusion if LtrA (see methods) to electroporate HEK 293 cells. Among the conditions tested, settings with voltage at 230 v and capacitance at 100 μF or 250 μF, labeled as (230 v, 100 μF) and (230 v, 250 μF) in Figure 5.4(a), respectively, gave best results when $2x10^6$ cells were mixed with 10 μg of RNP in 0.4 ml

of ice-cold PBS. Twenty-four hours post electroporation, the majority of the cells were transfected, as judged by fluorescence microscopy (not shown). An immunofluorescence assay using an anti-LtrA antibody on the same set of cells showed similar staining pattern (Figure 5.4(a)) with stronger fluorescence intensity because the secondary antibody was conjugated with fluorescein, which has the same spectrum as GFP. RNPs were in speckles in the cells, some in the cytoplasm and some in the nucleus. An NLS fusion on the LtrA protein in RNPs may direct more RNP particles into the nucleus. In addition, electroporated cells showed weak but uniform increase increase in fluorescence intensity compared with untransfected cells (Figure 5.4(b)).

5.3.2 Electroporation does not impair RNP activity

To find out whether RNPs would remain active after electroporation, I diluted 10 μg of anti-rDNA RNPs in 400 μL of ice-cold PBS and subjected them to electroporation conditions. RNPs were then concentrated to the original volume before dilution by using a Microcon column (Millipore) and compared in a TPRT assay, which uses ^{32}P-labeled dTTP to trace cDNA synthesis, with untreated RNPs and RNPs diluted and re-concentrated but not subjected to electroporaton. No difference in activity could be detected among the samples (not shown). Thus, electroporation conditions *per se* do not affect RNP activity.

5.4 *In vivo* homing by RNP electroporation

I electroporated HEK 293 cells and HEK 293 cells treated with sodium butyrate with RNPs against rDNA using similar conditions (230 v, 250 μF). The RNPs were purified from *E. coli*, and the LtrA protein in the RNP had a C-terminal fusion of the NLS of the HIV-1 Tat protein. Sodium butyrate is a histone deacetylase inhibitor, which keeps histones acetylated and helps relax the constraint of histone tails on chromosomal

DNA (Chen & Picaard, 1997). After electroporation, cells were seeded into growth medium and grown overnight. Twelve hours later, I added $MgCl_2$ and $MgSO_4$ to a final concentration of 10 mM each. Chromosomal DNA was prepared 72 hours after electroporation, and nested PCR was used to detect targeting events.

Figure 5.5 shows that the 3' junction of a targeted rDNA copy was amplified from electroporated 293 cells but not from electroporated 293 cells treated with sodium butyrate. The amplified band was gel-purified and sequenced to confirm the correct junction sequence. The 5' junction was not tested because of the interference of template switching (see section 5.1.2). Thus, RNPs are able to target chromosomal genes *in vivo*, and sodium butyrate did not help targeting. On the contrary, it seemed to prevent it from occurring. The *in vivo* genomic targeting is likely a very low-frequency event. However, improved *in vitro* preparations of highly concentrated RNPs may increase the targeting efficiency.

5.5 DISCUSSION

In this chapter, I used RNPs either purified from *E. coli* or reconstituted *in vitro* in both *in vitro* and *in vivo* experiments to target the chromosomal genes of human or mouse cells. The purpose of the *in vitro* experiments was to test whether the RNPs could find their targets in complex mammalian genomes and function under conditions in nuclear lysates without additional salt and Mg^{2+}.

Group II intron RNPs were able to target genes in purified chromosomal DNA and nuclear lysates *in vitro*. The *in vitro* experiments demonstrated that RNPs were able to identify and insert into their targets in the context of whole human and mouse genomes (Figures 5.2 and 5.3). However, *in vitro* targeting was not efficient, judged by the fact that even with rDNA (400 copies/cell), nested PCR was required to detect insertions. On the other hand, nested PCR also detected insertions into low copy number, even single-

copied, genes in purified chromosomal DNAs (Figure 5.2(b)). Targeting was successful in nuclear lysates albeit less efficient. When targeted in purified DNA, L1 insertion junctions were detected by one-step PCR, whereas in lysates, nested PCR was required (Figure 5.2(a), lane 3, and Figure 5.3, lanes 5 and 6). And nested PCR readily detected insertions into rDNA target in purified chromosomal DNAs but not in lysates (Figure 5.2(a), lane 1, and Figure 5.3, lane 1). However, Mg^{2+}-containing buffer was not necessary (Figure 5.3, lanes 5 and 6). Apparently, chromatin structures did not block the access of RNPs to chromosomal targets, and lysates could support the reaction without additional Mg^{2+}. However, it was not clear whether the insertion was complete because the PCR used to detect the 5' insertion junction was complicated by template switching. But it is reasonable that less 5' insertion junction is formed than the 3' junction. The formation of the 5' junction requires complete reverse transcription of the inserted intron RNA, whereas the 3' junction needs less than 200 bp of reverse transcription to be detected by the 3' end intron primers used in TPRT and PCR reactions. Incomplete reverse transcription could be due to either low processivity of the RT, or degradation of the inserted RNA template under the conditions in the experiments.

Injection and electroporation were the two major methods tested in this work to deliver RNPs into cells. In the mouse injection experiments, I could not detect any targeting event either on chromosomes or when target plasmid was co-injected with RNPs. The injection approach is largely limited by the relative low numbers (about 100) of eggs that could be injected in each experiment, compared to 10^6 to 10^7 cells used in electroporation experiments. For detection of a rare event, the difference in cell numbers could be critical. However, injections into *Xenopus* oocytes and zebra fish and *Drosophila* embryos were successful for plasmid targeting. The difference may be technical. The volume of RNPs that can be injected into mouse eggs is much smaller than

that for *Xenopus* oocytes, zebra fish or *Drosophila* embryos, and the RNP concentration needs to be below a certain level (about 1 μg/ml) to avoid clogging the smaller needles used for injecting mouse eggs. New methods of preparing RNPs with higer concentrations and solubility may improve the efficiency of injection.

I showed that RNPs could be incubated under electroporation conditions without losing activity. By using electroporation, I was able to detect an insertion event in 45S rDNA in HEK 293 cells. On one hand, targeting under current conditions is a rare event. Many factors could potentially affect the reactions. RNP preparations may differ in activity from batch to batch. Standardizing reaction conditions would be helpful.

In short, the group II intron Ll.LtrB was able to target mammalian genomes. It will be necessary to improve targeting efficiency so that group II introns would be generally useful for functional genomics and gene therapy.

5.6 SUMMARY

In this chapter, I explored the possibility of using RNPs to target mammalian chromosomal genes. First, I showed that RNPs targeted even a single-copy gene in purified mouse genomic DNA. Second, I demonstrated that RNPs were able to disrupt targets within chromatin structures in nuclear lysates, and nuclear lysates had sufficient Mg^{2+} and salt for targeting to occur. Third, I established that RNPs could be introduced into cells by electroporation, and I was able to detect a targeting event in the human rDNA gene. Increasing targeting efficiency will be necessary for this to be a generally useful procedure.

Figure 5.1. Retargeted introns used in this chapter and their mobility frequency in a plasmid assay in *E. coli.*

Each target is listed with the gene name and -1 position of the insertion site in the gene. s, sense strand insertion; a, anti-sense strand insertion. EBS and δ sequences of each intron are aligned to IBS and δ' sequences of the respective target. Base-paired nucleotides are connected with a vertical bar. Residues in IBS and δ' sequences of the target site are numbered according to Guo et al. (2000). Nucleotides at positions -12, -7, +1, and +5 are in bold. The plasmid-base mobility assay was done by cotransforming a donor plasmid expressing the intron RNA and IEP and a recipient plasmid carrying a promoterless tetracycline-resistance gene proceeded by a specific target site. The intron RNA carries a T7 promoter in DIV, which will result in tet^R expression upon intron insertion into the target site. The mobility frequency refers to percentage of tetracycline-resistant cells among all transformed cells. The mobility frequencies of retargeted introns were normalized against that of the wild-type intron.

Figure 5.1

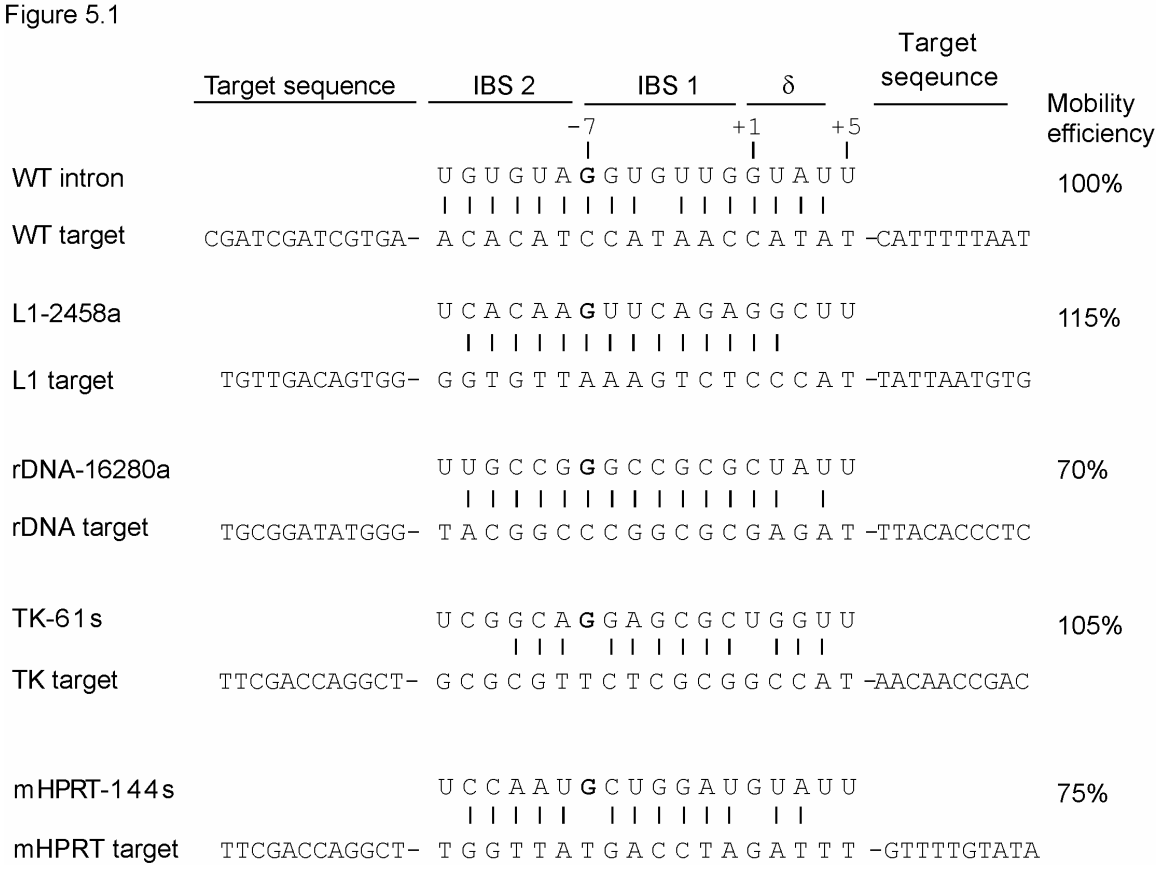
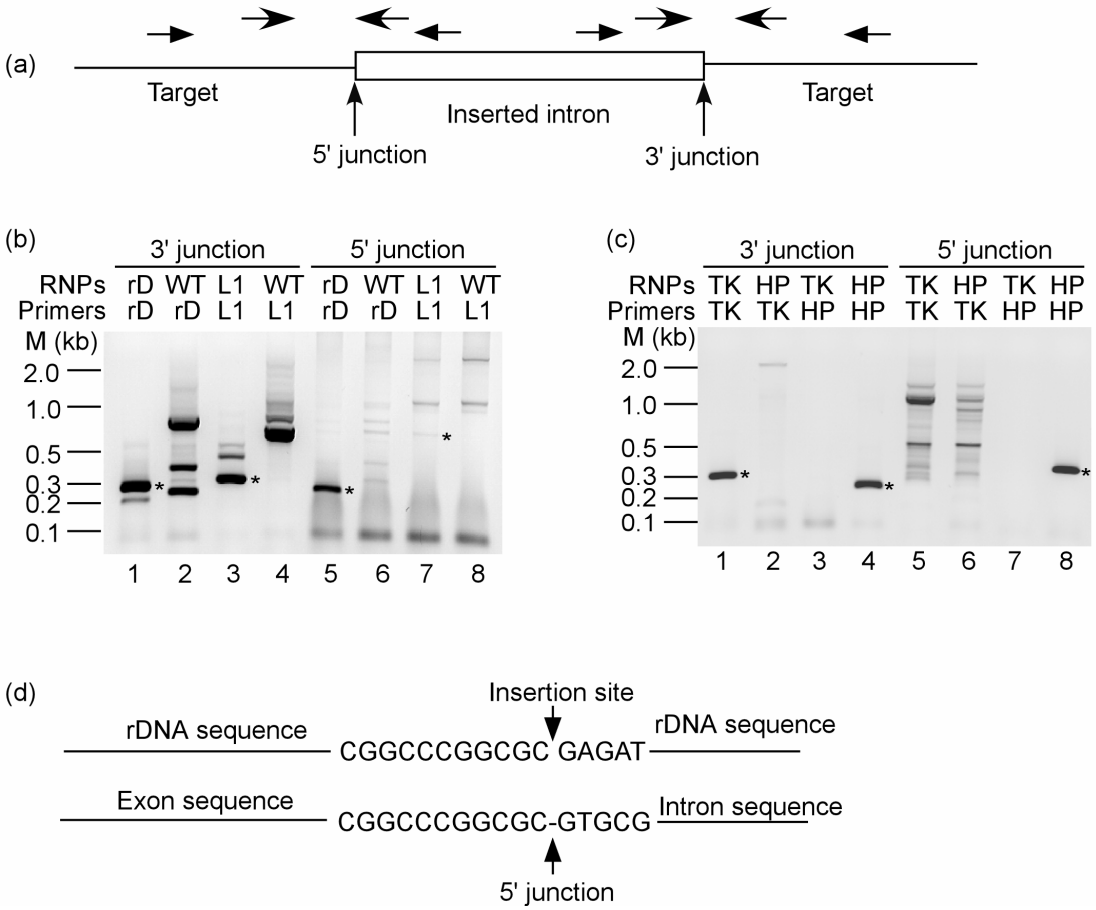

Target sequence
IBS 2
IBS 1
δ
Target seqeunce
Mobility efficiency
-7
+1
+5
WT intron
U G U G U A G G U G U U G G U A U U
100%
WT target
CGATCGATCGTGA- A C A C A T C C A T A A C C A T A T -CATTTTTAAT
L1-2458a
U C A C A A G U U C A G A G G C U U
115%
L1 target
TGTTGACAGTGG- G G T G T T A A A G T C T C C C A T -TATTAATGTG
rDNA-16280a
U U G C C G G G C C G C G C U A U U
70%
rDNA target
TGCGGATATGGG- T A C G G C C C G G C G C G A G A T -TTACACCCTC
TK-61s
U C G G C A G G A G C G C U G G U U
105%
TK target
TTCGACCAGGCT- G C G C G T T C T C G C G G C C A T -AACAACCGAC
mHPRT-144s
U C C A A U G C U G G A U G U A U U
75%
mHPRT target
TTCGACCAGGCT- T G G T T A T G A C C T A G A T T T -GTTTTGTATA

Figure 5.2 RNPs were able to insert into target sites in purified chromosomal DNAs.

(a) Schematic showing PCR amplification of intron/target site junction following a TPRT reaction. In each PCR reaction, one intron primer and one target primer were used. In nested PCR, two sets of intron/target primers were used in consecutive PCR reactions. (b) PCR of *in vitro* targeting of rDNA and L1 retrotransposon in HEK 293 DNA. RNP and primers used are indicated at the top of each lane. rD, rDNA; L1, the L1 retrotransposon; WT, the wild-type Ll.LtrB intron RNP. Lanes 1, 2, 5, and 6 show products of nested PCRs. The remaining lanes show products of one-step PCRs. (c) Nested PCR of *in vitro* targeting of the mouse HPRT gene and the HSV-TK gene. As in (b), RNPs and primers are labeled above the lanes. HP, mHPRT; TK, HSV-TK. (d) Schematic showing the GC-rich homologous region between the rDNA donor plasmid and target site, resulting in template switching detected in nested PCR. An asterisk was placed to the right of each correct PCR band.

Figure 5.2


(a)
Target
Inserted intron
Target
5' junction
3' junction
(b)
3' junction
5' junction
RNPs rD WT L1 WT rD WT L1 WT
Primers rD rD L1 L1 rD rD L1 L1
M (kb)
2.0
1.0
0.5
0.3
0.2
0.1
1 2 3 4 5 6 7 8
(c)
3' junction
5' junction
RNPs TK HP TK HP TK HP TK HP
Primers TK TK HP HP TK TK HP HP
M (kb)
2.0
1.0
0.5
0.3
0.2
0.1
1 2 3 4 5 6 7 8
(d)
Insertion site
rDNA sequence
CGGCCCGGCGC GAGAT
rDNA sequence
Exon sequence
CGGCCCGGCGC-GTGCG
Intron sequence
5' junction

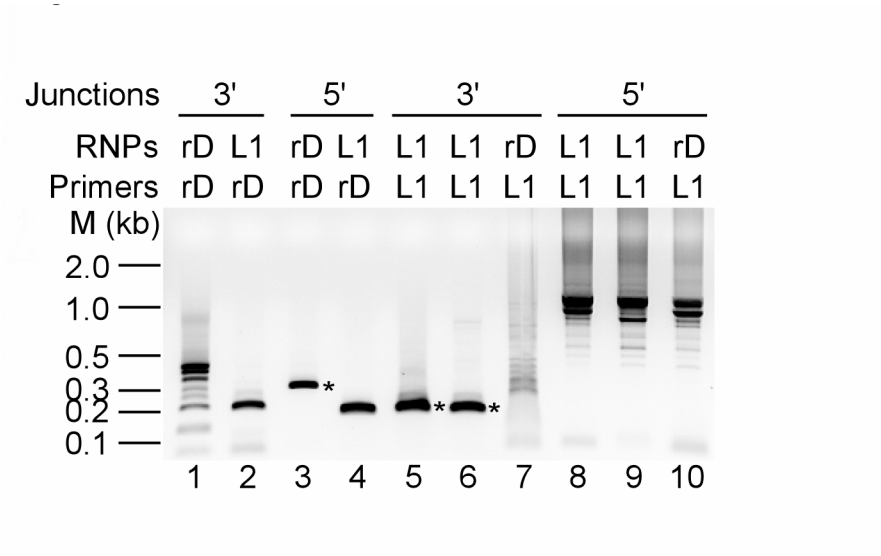


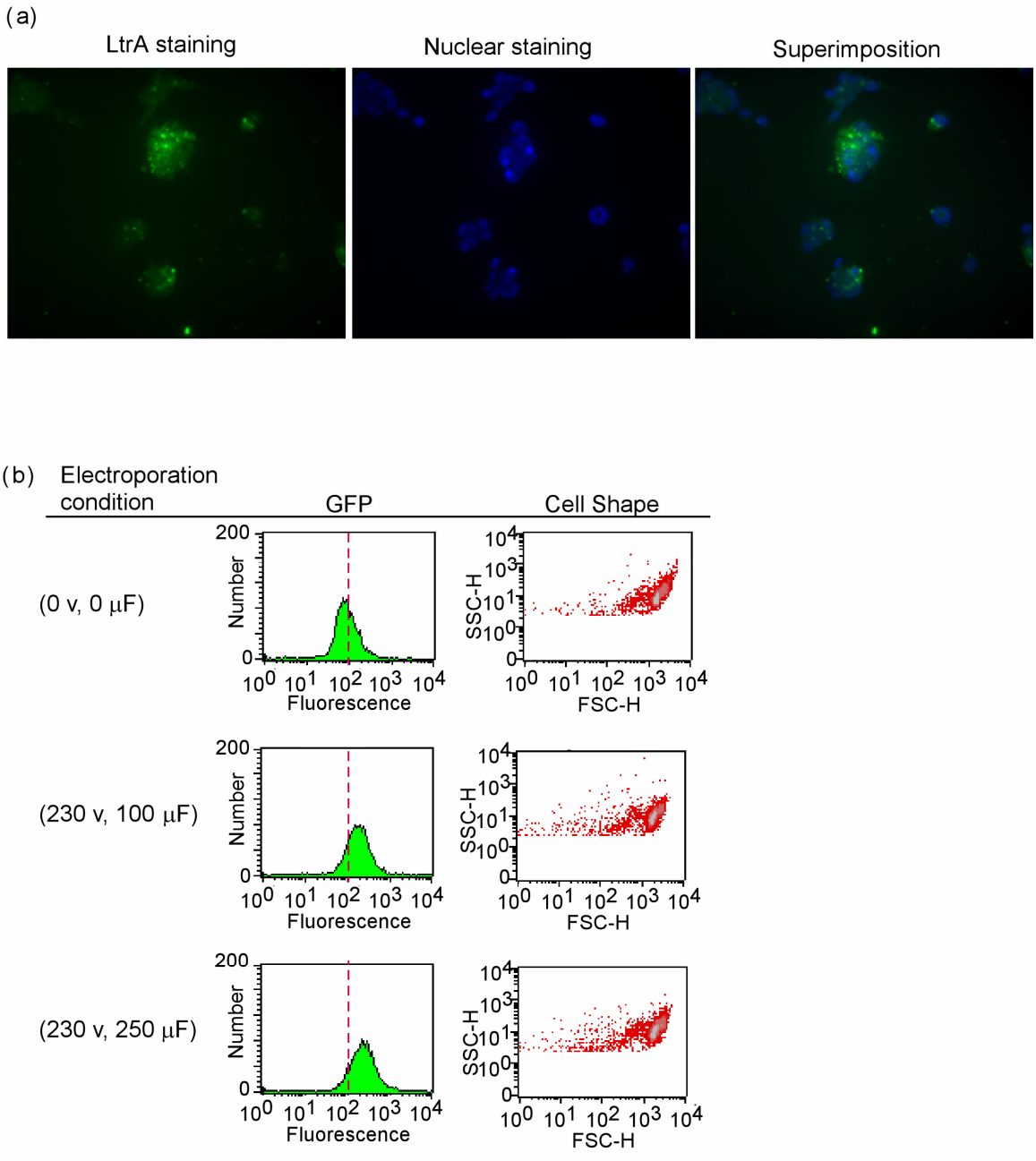
Figure 5.3 Nested PCR showing L1 interruption by specific RNPs in nuclear lysates.

TPRT reactions were done with or without added Mg^{2+}-containing buffer (40 mM Tris-HCl, pH 7.5, 10 mM KCl, 10 mM $MgCl_2$, 5 mM DTT) for L1 targeting. Lanes 5 and 8 were with added buffer, and lanes 6 and 9 were without the buffer. All contained 2 mM dNTPs. Each RNP was used as a negative control for the other target site (lanes 2, 4, 7, and 10). The single band in lane 2 was sequenced and shown to result from nonspecific amplification of the rDNA. The bands in lanes 5 and 6 were confirmed to have the correct 3' integration junction for L1 by sequencing. All the bands in lane 1 resulted in nonspecific amplification. Each correct PCR product is indicated by an asterisk.

Figure 5.4 Electroporation of RNPs into cells.

(a) GFP fused to the N-terminal of the LtrA protein was visualized with a fluorescence microscope (left panel). In a superimposition to the nuclear staining (middle panel), GFP is shown to be in or around the nuclei. (b) Cells from (a) were trypsinized and measured with a Fluorescence-activated Cell Sorter (FACS). Electroporation conditions are indicated on the left. The histograms on the left show the shift in fluorescence intensity upon electroporation of GFP-containing RNP complex (with filter FL1, 530±30 nm). The histograms on the right show that cell shape did not change after electroporation.

Figure 5.4


(a)
LtrA staining
Nuclear staining
Superimposition
(b)
Electroporation condition
GFP
Cell Shape
(0 v, 0 μF)
(230 v, 100 μF)
(230 v, 250 μF)
Number
Fluorescence
SSC-H
FSC-H

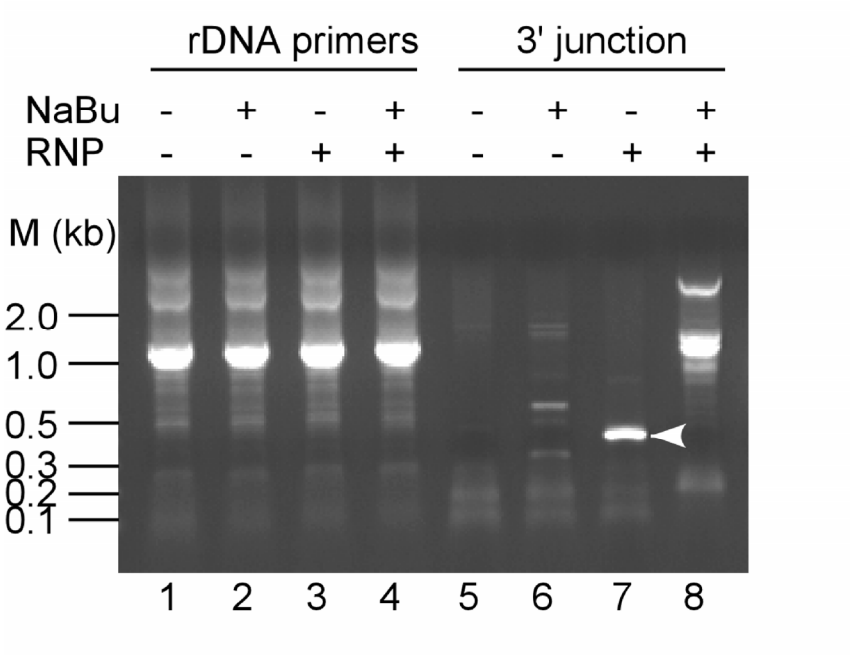


Figure 5.5 *In vivo* targeting of the rDNA gene in HEK 293 cells.

HEK 293 cells and sodium butyrate (NaBu, 3 mM)-treated HEK 293 cells were electroporated (230 v, 250 μF, $2x10^6$ cells in 400 μl PBS) with RNPs against rDNA. Chromosomal DNAs were purified (see methods) 72 h after electroporation. Nested PCR were done to detect insertion junctions. Lanes 1-4 show PCR products obtained using rDNA primers to detect rDNA in the purified chromosomal DNA. Lanes 5-8 were amplified as shown in Figure 5.2(a). The specific band in lane 7 (a white arrowhead on the right) was confirmed as correct junction by sequencing.

Chapter 6 Materials and methods

6.1 MATERIALS AND METHODS FOR STUDYING LTRA-PROMOTED SPLICING OF THE LL.LTRB INTRON IN *E. COLI*

6.1.1 *E. coli* strains and growth conditions

E. coli HMS174(DE3) (Novagen, Madison, WI) was used for *in vivo* splicing assays; BL21(DE3) and Rosetta (DE3) (Novagen) were used for LtrA protein expression; and DH10B and DH5α were used for cloning. Bacteria were grown in LB medium. Ampicillin and chloramphenicol were added at 100 μg/ml and 25 μg/ml, respectively.

6.1.2 Recombinant plasmids

Plasmids used for *in vivo* splicing assays contain *ltrB*/GFP fusions cloned downstream of the phage T7 promoter and ϕ10 gene Shine-Dalgarno sequence in the vectors pAT or pET, which are derivatives of pACYC184 and pET-11a (New England Biolabs, Beverly, MA), respectively (see below). The *ltrB*/GFP fusions consist of the first eight codons of ϕ10 linked in-frame to a segment of the *L. lactis ltrB* gene (58 bp of exon 1 (E1), the Ll.LtrB intron, and 38 bp of exon 2 (E2)), with E2 linked in frame to codons 2 to 238 of the GFP ORF. The GFP ORF used here was derived from that in pGFPuv (Clontech, Palo Alto, CA) by introducing amino acid substitutions F64L and S65T to improve performance in FACS assays (Cormack et al., 1996). pAT and pET were derived from their parent vectors by inserting *E. coli rrnB* T1 and T2 transcription terminators to mitigate the effect of strong T7 RNA polymerase transcription on plasmid stability. pAT contains a T1 terminator (250-bp *Aat*II/*Sal*I fragment of pKK232-8; Amersham Pharmacia; Piscataway, NJ) inserted between the *Nru*I and *Sal*I sites of pACYC184, and T1 + T2 terminators (0.9-kb PCR product of pKK232-8 synthesized with primers T1T2-

5'ClaI (5’-GAACATATCGATAGCGTCCGCCATC) and T1T2-3'XbaI (5’-AGCACGATCTAGAGCACCCGTGGC) inserted between the pACYC184 *Cla*I and *Xba*I sites. pET contains T1 and T2 terminators (960-bp *Cla*I/*Ssp*I fragment of pAT) inserted between the *Cla*I and *Sal*I sites of pET11a.

Plasmids in the pALG1 series contain *ltrB/*GFP fusions with the wild-type (2.6-kb) Ll.LtrB intron or its mutant derivatives cloned in pAT. pALG1-WT contains the wild-type Ll.LtrB intron recloned from pLI1 (Matsuura et al., 1997). pALG1-ΔORF contains a mutant intron deleted for most of the intron ORF (positions 691-2286) (Matsuura et al., 1997), and pALG1-ΔDV contains a mutant intron deleted for DV (positions 2398-2492) (Guo et al., 2000). The 0.9-kb ΔORF intron in pALG1-ΔORF has the same sequence at the wild-type intron, except for an *Mlu*I site inserted in domain IV during construction of the deletion.

Plasmids in the pALG2 series contain *ltrB/*GFP fusions with the 0.9-kb Ll.LtrB-ΔORF intron and have the LtrA protein expressed from the downstream *cis* position. pALG2-WT was derived from pALG1-ΔORF by inserting a 1.8-kb PCR product containing the LtrA ORF (amplified from pLI1 (see below) with primers 5’-GGACTAGTGATTAGGGAGGAAACCTC and 5’-GGG-GGCCCGTTCGTTCGTAAAAATTCACTTG) between the *Spe*I and *Apa*I sites of pALG1-ΔORF. pALG2 derivatives with deletions of various intron domains were constructed by swapping in the 0.9-kb *Pvu*I/*Kpn*I fragments of previously constructed pGM-ΔORF-based plasmids containing the deletions (Wank et al., 1999).

pELG1-ΔORF contains a GFP fusion with the 0.9-kb Ll.LtrB-ΔORF intron constructed by ligating the 1.9-kb *Bgl*II/*Cla*I fragment of pALG1-ΔORF between the corresponding sites of pET. This plasmid was used for the insertion of LtrA mutant libraries at the downstream *cis* position for unigenic evolution analysis (see below).

pELG2-WT contains the same GFP fusion with the 0.9-kb Ll.LtrB-ΔORF intron and the LtrA protein expressed from the downstream *cis* position. It was constructed by simultaneously inserting the 1.9-kb *Bgl*II/*Spe*I fragment containing the *ltrB*/GFP fusion of pALG1-ΔORF and a 1.8-kb PCR product containing the LtrA ORF [amplified from pLI1 with primers LtrA 3'-ClaI (5'-GGGGGCATCGATGCCCTTCGTTCGTAAAAATTCA) and LtrA 5'-SpeI (5'-GGGACTAGTACTTTAAGAAGGAGATATACATATG) and digested with *Cla*I and *Spe*I] between the *Cla*I and *Bgl*II sites of pET.

Two related constructs were used to express the LtrA protein in *trans*: pLI1P, which contains the LtrA ORF cloned downstream of the phage T7 promoter in pET-11a (Matsuura et al., 1997) and pET-LtrA, which contains the LtrA ORF cloned downstream of the phage T7 promoter in pET. pET-LtrA was derived from pELG2-WT by replacing a 1.9-kb *Bgl*II/*Spe*I fragment containing the T7-ltrB/GFP fusion with a double-stranded DNA oligonucleotide containing the T7 promoter flanked by *Bgl*II and *Spe*I sites. Both constructs gave splicing efficiencies similar to that of pALG1-ΔORF in FACS assays, but pLI1P contains additional DIV[3']-E2 sequences, which prevents its use in RT-PCR and primer extension assays.

pALE contains the ligated *ltrB* exon sequence (132-bp RT-PCR product amplified from spliced *ltrB* RNA with primers 5'-CCAAGCTTAGGTGGCGAATATGAATTTGTG and CCCTGCAGTTGCCAGTATAAAGATTCGTAG) cloned downstream of the T7 promoter between the *Hin*dIII and *Pst*I sites of pALG1-ΔORF.

pGMΔORF, used to synthesize *in vitro* transcript substrates for biochemical assays, contains a 0.9-kb Ll.LtrB-ΔORF intron, cloned behind the phage T3 promoter in pBSKS+ (Stratagene, LaJolla, CA) (Saldanha et al., 1999). This plasmid has

compensatory base changes (U208C, A341G) at two positions in the intron's DId(iii) stem that introduce a *Hin*dIII site to facilitate mutagenesis (Saldanha et al., 1999). These changes were confirmed to have no effect on the rate of *in vitro* splicing at 1:1 molar ratio of LtrA protein.

pELG1-ΔDIVa and pELG2-ΔDIVa were constructed from pELG1-ΔORF and pELG2-ΔORF, respectively, by swapping in the 0.9-kb *Pvu*I/*Kpn*I fragments of the previously constructed plasmid pGM-ΔORF/ΔDIVa (Wank et al., 1999).

LtrA mutants were constructed as described previously via a two-step PCR with Vent polymerase, with the desired modifications introduced via one of the primers (San Filippo et al., 2002). The PCR products containing the mutations were digested with restriction enzymes and swapped for the corresponding segments of the LtrA ORF in either pELG2-ΔORF, pELG2-ΔDIVa, or the intein-based expression vectors pImp-1P (Saldanha et al., 1999) or pImp-10P. The latter is a derivative of pImp-1P with silent mutations that introduce restriction sites *Eag*I (positions 746-751), *Sac*I (positions 1333-1338), *Bgl*II (positions 1666-1671), *Sac*II (positions 1880-1885), *Bam*HI (positions 2042-2047), and *Aat*II (positions 2213-2218), thereby facilitating the construction of mutant proteins. Regions that were generated by PCR were sequenced to confirm the introduction of the mutation and absence of any adventitious mutations.

6.1.3 FACS assay of maturase-promoted group II intron splicing

A single *E. coli* colony freshly transformed with a pAT- or pET-based plasmid containing an *ltrB*/GFP fusion was inoculated into 1 ml of LB medium containing ampicillin and grown overnight at 37°C. Fifty µl of the overnight culture was then inoculated into 1 ml of fresh medium, grown with shaking at 37°C to $O.D._{595} = 0.4$, and induced with 0.1 mM IPTG for 3 h. Cells were collected by centrifugation, resuspended in phosphate-buffered saline (PBS, 140 mM NaCl, 2.7 mM KCl, 9 mM Na_2HPO_4, 1.6

mM KH_2PO_4, pH 7.4) and analyzed using a FACS Caliber (Becton Dickinson, Immunocytometry Systems, San Jose, CA), with filter FL1 (530 ± 30 nm). The results were analyzed with the CELLQuest program (Becton Dickinson).

6.1.4 Immunoblotting

Cells were grown and induced with IPTG, as described above for FACS assays. To prepare total protein for SDS-PAGE, half of the cells were pelleted by centrifugation, dissolved in 1X SDS-PAGE loading buffer, boiled for 3 min, and microfuged for 5 min to pellet any residual insoluble material. Soluble protein was prepared by incubating the other half of the cells with lysozyme (400 μg/ml) for 15 min on ice, followed by three cycles of freezing-thawing between -80°C and room temperature, and then microfuging for 5 min to pellet insoluble material. Total and soluble proteins were electrophoresed in an 8% polyacrylamide/1% SDS gel, which was transferred to a nitrocellulose membrane with a Hoefer SemiPhor blotter (Amersham Pharmacia Biotech, Piscataway, NJ). The membrane was blocked with 3% dry milk (Nestle USA Inc., Solon, OH) and 1% bovine serum albumin (Sigma, St. Louis, MO) in TTBS (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2.5 mM KCl, 0.5% Tween 20) for at least 1 h at room temperature, and then incubated for 1 to 2 h with a 1:1000 dilution of a rabbit polyclonal anti-LtrA antibody preparation (obtained from Dr. Gary Dunny, Univ. of Minnesota). The blots were washed with TTBS four times at room temperature for five min each time, then incubated with goat anti-rabbit secondary antibody (Pierce, Rockford, IL) at 1:100,000 dilution for 1 h at room temperature, washed another four times in TTBS, and developed with SuperSignal West Pico Chemiluminescent substrate (Pierce). After development, the blots were stained with AuroDye Forte (Amersham Biosciences) to confirm equal loading.

6.1. 5 RT-PCR assay of *in vivo* RNA splicing

RT-PCR was carried out on total cellular RNA, which was isolated as described (Belfort et al., 1990) and further purified with a Qiagen RNeasy mini kit (Qiagen, Inc., Valencia, CA). For the initial reverse transcription reaction, 5 μg of RNA was mixed with 1 pmole of primer PPI (5'-GTTCTTCTCCTTTACTC) in 50 mM Tris-HCl, pH 7.5, 60 mM NaCl, 10 mM DTT and annealed by heating to 60°C for 3 min and then chilling on ice. After additions to give 1.8 mM $MgCl_2$ and 125 μM of each dNTP, the reaction was initiated by adding M-MLV RT (6 units; Invitrogen, Carlsbad, CA), incubated at 48°C for 30 min, and terminated by adding EDTA to 10 mM. 1/20 of the product was then used as a template for PCR in 50 μl of reaction medium containing 2.5 units Taq DNA polymerase (Invitrogen), 150 μM dNTPs, 3.5 mM $MgCl_2$, 1X PCR buffer, and 25 pmoles each of primers PPI (see above) and PHIII (5'-CATATGGCTAGCATGACTGG) for 16 cycles of 94°C for 30 sec, 50°C for 30 sec, and 72°C for 1.5 min. The products were analyzed in a 1.0% agarose gel containing ethidium bromide (0.5 μg/ml).

6.1.6 Poisoned primer extension assay of *in vivo* splicing

Cellular RNA (5 μg), isolated as described above, was annealed to 5' ^{32}P-labeled primer PRET (5'-CGTAGAATTAAAAATGATATG; 0.2 pmole, ~1.3 x 10^7 cpm) by incubating at 60°C for 3 min in 7 μl of 50 mM Tris-HCl, pH 8.0, 60 mM NaCl, 10 mM DTT and then chilling on ice. Reverse transcription reactions were as described above for RT-PCR assays, except that 500 μM ddCTP was used instead of dCTP. The reaction was terminated by adding an equal volume of gel loading buffer (80% formamide, 10 mM EDTA, pH 8.0, 0.025% bromophenol blue, and 0.025% xylene cyanol FF), and the products were analyzed in a denaturing 10% polyacrylamide gel, which was dried and quantified with a PhosphorImager.

6.1.7 Colony-based fluorescence assays

Cells were plated on LB containing ampicillin at a density of 1200-1500 colonies per 150 mm diameter plate and incubated overnight at 37°C. The colonies on each plate were then lifted onto five to seven 35-mm diameter nitrocellulose filter circles, and the filters were incubated on petripads (Millipore, Billerica, MA) saturated with LB medium containing ampicillin and IPTG (0.2 mM) for 4 h at 37°C. For fluorescence assays, the filters were scanned with a FluorImager SI (Molecular Dynamics, Sunnyvale, CA), using an excitation wavelength of 488 nm and a built-in 515-nm long-pass emission filter. Splicing-competent colonies that produce GFP appeared as black dots against a light background.

6.1.8 Colony western assay

Colonies were plated, transferred to nitrocellulose filters, and induced with IPTG as described above for the colony-based fluorescence assay, then lysed by incubating filters on drops of protoplast buffer (15 mM Tris-HCl, pH 8.0, 0.45 M sucrose, 8 mM EDTA, 0.4 mg.ml lysozyme) for 15 min at 4°C, followed by lysis buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 1 mM sodium citrate, 1.5% SDS) for 10 min at room temperature. The filters were then washed twice in the DNase I Buffer (40 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$) at room temperature and incubated with drops containing 50 units/ml DNase I in buffer at 37°C for 15 min. For immunoblotting, the filters were first blocked by incubating in TTBS containing 5% dry milk and 1% BSA at room temperature for 1 h. Cell debris was wiped away with a Kimwipe. The filters were incubated with primary anti-GFP antibody (Clontech antibody; 1:3000 dilution) for 2 h at room temperature, washed four times with TTBS at room temperature for 5 min each time, and then incubated with HRP-conjugated secondary antibody for 1 h at room temperature. After several washes, the filters were developed using the supersignal West

Dura Extended Duration substrate (Pierce) and exposed to X-ray films. Both the primary and secondary antibodies had been pre-adsorbed with pET2-transformed cell lysate at room temperature for 0.5 h. Splicing-competent colonies were identified by aligning the exposed and developed X-ray film and corresponding filters. Splicing was confirmed by RT-PCR.

6.1.9 Unigenic evolution

For the unigenic evolution analysis in chapter 2, a library of LtrA proteins with random mutations was constructed by error-prone PCR (Fromant et al., 1995) in 100 µl of reaction medium containing 0.06 pmol of pLI1 template, 50 pmoles each of primers LtrA 5'-SpeI and LtrA 3'-ClaI (see above), 0.39 mM dATP, 0.15 mM dCTP, 1.17 mM dGTP, 3.85 mM dTTP, 11 mM $MgCl_2$, 0.5 mM $MnCl_2$, and 5 units of Taq DNA polymerase (Invitrogen). The reaction conditions were initial denaturation at 94°C for 5 min, followed by 16 cycles of 91°C for 1 min, 51°C for 1 min, and 72˚C for 10 min, with final extension at 72°C for 15 min. The 1.8-kb PCR product containing the mutagenized LtrA ORF was then gel-purified, digested with *Spe*I and *Cla*I, and cloned between the corresponding sites of pELG1-ΔORF. The mutation frequency determined by sequencing 36 randomly selected clones from the library was 0.78%, corresponding to ~10 amino acid substitutions per protein.

After transforming the library into HMS174(DE3), splicing-competent LtrA variants were identified by colony-based fluorescence assay (see below). Individual colonies were then grown up in LB medium containing ampicillin, and plasmids encoding LtrA variants were isolated and transformed into HMS174(DE3) to retest their splicing phenotype by FACS assays. Variants that spliced the Ll.LtrB-ΔORF intron with efficiency ≥ 50% that of wild-type LtrA assayed in the same experiment were sequenced to identify changes in the LtrA ORF.

Mutational data were analyzed as described by Deminoff *et al* (1995). The mutability value (M) in a 25-amino acid sliding window was calculated by using the formula:

$$M = \frac{f_{Omis}}{f_{Emis}} - 1,$$

where f_{Omis} is the observed frequency of missense mutations and f_{Emis} is the expected frequency. The latter was calculated for each codon based on the probability of a single nucleotide change producing a missense mutation, correcting for the transition/transversion ratio of 3.4 for the variants selected from the library. Negative mutability values indicate hypomutability, with a maximum value of -1 indicating no missense mutations in a given window. Positive values indicate hypermutability and were normalized by using the formula:

$$\frac{1}{f_{Emis}} - 1,$$

so that all mutability values fall between -1 and +1. Chi square was calculated for each hypomutable region by using the equation:

$$\chi^2 = \frac{(abs(O_{mis} - E_{mis}) - 0.5)^2}{E_{mis}} + \frac{(abs(O_{sil} - E_{sil}) - 0.5)^2}{E_{sil}},$$

where "abs" indicates the absolute value; O_{mis} and O_{sil} are numbers of missense and silent mutations observed in a given region; and E_{mis} and E_{sil} are expected numbers of missense and silent mutations in the region. Boundaries were moved inward to improve chi square scores when necessary. Hypomutable regions are defined as the longest continuous stretch of amino acid residues that has a P-value < 0.01.

The LtrA library used in chapter 3 was generated in the same way by error-prone PCR and cloned into *Spe*I and *Cla*I sites of pELG1-ΔDIVa. Based on sequencing results of 48 randomly picked clones, the mutation rate of the library was 0.6%. The ratio for transition to transversion was 3.0. On average, each mutant contained 10 mutations.

Positive clones were identified by using the Colony Western assay (see below). Mutational data were analyzed as described above except that positive mutability values were not normalized.

6.1.10 Biochemical assays

Wild-type and mutant LtrA proteins were expressed from the intein-based expression plasmid pImp1P or its mutant derivatives in *E. coli* BL21(DE3) or Rosetta (DE3) and purified on a chitin-affinity column, as described (Saldanha et al., 1999). Protein concentrations were measured by Bradford assay against a standardized LtrA preparation, whose concentration had been determined spectrophotometrically based on its calculated extinction coefficient (Saldanha et al., 1999). The wild-type and mutant protein preparations were > 99% and 80-99% pure, respectively, as judged by Coomassie-blue stained SDS-gels.

RT assays with poly(rA)/oligo$(dT)_{18}$ were carried out with 20 nM LtrA protein and 1 μg substrate per 10 μl reaction, as described (Matsuura et al., 1997). RT assays with Ll.LtrB/E2+10 were carried out in 10 μl of reaction medium containing 20 nM protein, 40 nM Ll.LtrB template, 400 nM E2+10 primer, 450 mM NaCl, 5 mM $MgCl_2$, 40 mM Tris-HCl, pH 7.5 plus 10 μCi [α-^{32}P]dTTP (3,000 Ci/mmol; New England Nuclear, Boston, MA) and 0.2 mM each of dATP, dCTP, and dGTP (Wank et al., 1999). Reactions were initiated by addition of dNTPs and incubated at 30°C for 30 min, which was shown previously to be in the linear range. In both RT assays, incorporation of [α-^{32}P]dTTP into high molecular weight material was measured by spotting onto DE81 paper (Whatman, Maidstone, England) and counting Cherenkov radioactivity in a scintillation counter (LS6500, Beckman, Fullerton, CA).

RNA splicing and binding reactions were carried out in reaction medium containing 450 mM NaCl and 5 mM $MgCl_2$, the relatively high salt concentration being

necessary for specific binding and optimal activity of the LtrA protein (Saldanha et al., 1999). For RNA splicing assays, ^{32}P-labeled RNA substrates were synthesized by *in vitro* transcription of *Bam*HI-linearized pGMΔORF with phage T3 RNA polymerase (Stratagene) and purified in a denaturing 4% polyacrylamide gel (Wank et al., 1999). *In vitro* splicing reactions were carried out with ^{32}P-labeled gel-purified RNA substrate (20 nM; 2-4 x 10^4 cpm) and wild-type or mutant LtrA protein (20 or 200 nM) in 150 µl of 450 mM NaCl, 5 mM $MgCl_2$, 40 mM Tris-HCl, pH 7.5, 5 mM dithiothreitol (DTT) and 1 unit/µl RNasin (Wank et al., 1999). Splicing reactions were initiated by adding LtrA protein and incubated at 30^{o}C. At different times, 5 to 10 µl portions were removed, and the reaction was terminated by extraction with phenol-chloroform-isoamyl alcohol (25:24:1). After centrifugation, the aqueous phase was mixed with an equal volume of loading dye, and products were analyzed in a denaturing 4% polyacrylamide gel, which was dried and quantified with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). Data from splicing time courses were best fit to single or double exponential equations using KaleidaGraph (Synergy Software, Reading, PA) to obtain k_{obs}.

For equilibrium binding assays, ^{32}P-labeled RNA (5 pM) was incubated with increasing concentrations of LtrA protein in 500 µl of reaction medium containing 450 mM NaCl, 5 mM $MgCl_2$, and 40 mM Tris-HCl, pH 7.5. After incubation for 1 h at 30°C, RNA-protein complexes were isolated by filtration through a nitrocellulose filter (PROTRAN BA85, Scheicher & Schuell, Keene, NH). The filters were washed three times with 1 ml of reaction medium, dried, and counted for Cerenkov radioactivity. Varying the incubation time from 5 to 60 min had no significant effect on the binding curves.

k_{off} values were measured as described (Saldanha et al., 1999). Complexes were formed by incubating 20 nM LtrA protein with 20 nM ^{32}P-labeled Ll.LtrB RNA (2-5 x

10^5 cpm) in 100 µl of 450 mM NaCl, 5 mM $MgCl_2$, 40 mM Tris-HCl, pH 7.5, 5 mM DTT, 1 unit/µl RNasin, and 100 µg/ml bovine serum albumin. The complexes were then diluted 20-fold into the same reaction medium containing a 100-fold excess of unlabeled Ll.LtrB RNA, and the dissociation of the complexes as a function of time was followed by nitrocellulose filter binding. Unless noted otherwise, the data were best fit to equations with two exponentials reflecting rapidly dissociating and stable complexes, with the k_{off} values reported in the text being those for the stable complex (Saldanha et al., 1999).

6.2 MATERIALS AND METHODS FOR STUDYING SPLICING AND TARGETING OF THE LL.LTRB INTRON IN MAMMALIAN CELLS

6.2.1 *E. coli* strain, cell lines, and medium

Cell lines HEK 293 and COS-7 were obtained from Dr. Robert Krug. DMEM was from Invitrogen. Leibovitz's L-15 modified medium was from ATCC. Medium supplements were from Invitrogen. Fetal bovine serum was from Gemini. Cells were grown in a 37°C incubator with 5% CO_2 except when grown in Leibovitz's medium, where cells were kept airtight.

Transfection reagents used were Fugene 6 (Roche Applied Science, Indianapolis, IN), Lipofectamine 2000 (Invitrogen Corporation, Carlsbad, CA), and Polyfect (Qiagen Inc., Valencia, CA).

6.2.2 Recombinant plasmids

pCMV/myc/nuc was purchased from Invitrogen. pIRES was from Clontech. pCMV-LtrA was constructed by inserting the PCR amplified LtrA ORF into the *Nco*I and *Xho*I sites of pCMV/myc/nuc. pCMV-IVS was constructed by inserting the 214-bp *Alu*I fragment from pIRES into the *Pml*I site of pCMV/myc/nuc. pCMV-IVS-LtrA was constructed by replacing the *Nco*I and *Xba*I fragment of pCMV-IVS with the *Nco*I and *Xba*I fragment of pCMV-LtrA. pHHWT was constructed by inserting PCR-amplified

intron with flanking sequences into the two *Bsm*BI sites of pHH21 plasmid (Neumann *et al.*, 1999). pGMΔORF/NL was constructed by inserting the *Sal*I/*Xho*I fragment from pcDNA3.1/NL to the *Sal*I site of pGMΔORF. pcDNA3.1/NL-ΔORF was constructed by inserting the *Apa*I/blunted + *Xba*I fragment of pGMΔORF/NL into the *Eco*RV and *Xba*I sites of pcDNA3.1/NL-ΔORF. pTZU6+27-ΔORF was constructed by inserting the *Xho*I/*Xba*I fragment of pGMΔORF/NL into the *Sal*I and *Xba*I sites of pTZU6+27 (kindly provided by Engleke, U. of Michigan, Ann Arbor, MI). pBSΔORF.500 was constructed by replacing the fragment between the *Bse*RI and *Stu*I sites on pBSU3hTR500 with blunt-ended *Xho*I/*Bam*HI fragment from pGMΔORF. pBSU3ΔORFNL was constructed by replacing the *Bse*RI/*Eco*RV fragment with the blunt-ended *Afl*II/*Pvu*II fragment.

6.2.3 Codon optimization and recombinant plasmids

All oligonucleotides were synthesized by HHMI/Keck Oligonucleotide Synthesis Facility (Yale U., New Haven, CN). For segment 1, long oligonucleotides are hLtrAF1: 5'CCCACCATGGCCATCCTGGAGCGCATCAGCAAGAACTCACAGGAGAACAT-CGACGAGGTGTTCACCCGCCTGTACCGCTACCTGCTGCGCCCCGACATCTAC-TACGTGGCCTACC3'; hLtrAR1: 5'CTTCTCCTCGCTGAAGCCGTCGGCGGTGTC-GTCCAGGATGCCCTTGGTGCTGGCGCCCTTGTTGCTGTACAGGTTCTGGTAG-GCCACGTAGTAGATGTCGGGGC3'; hLtrAR2: 5'GATCCTCACGGCCTCCTGGAT-CAGCTTGTCGGTGAAGGTGGGGATGCCCAGGGGGCGCATCTTCTTGCTGTTC-TTCTTGGCGATGTACATGCGGCGCACGG3'; hLtrAF2: 5'CCGCCGACGGCTTCA-GCGAGGAGAAGATCAAGAAGATCATCCAGAGCCTGAAGGACGGCACCTACT-ACCCCCAGCCCGTGCGCCGCATGTACATCGCC3'; Short oligonucleotides are s-EcoRI: 5'AAAAGAATTCCACCATGAAGCCCCACCATGGCCATCCTGGAGC3' and s-KpnI: 5'AAAAGGTACCATGGACTCCAGGATGATCCTCACGGCCTCCTGGATC-AGC3'.

For segment 2, long oligonucleotides are hLtrAF3: 5'GTTCGAGGACGTGAGC-CACGGCTTCCGCCCCCAGCGCAGCTGCCACACCGCCCTGAAGACCATCAAGC GCGAGTTCGGCGGCGCCAGGTGGTTCGTGGAGGGCG3'; hLtrA-F4: 5'GATCAA-CCTGAAGATCAAGGACATGAAGATGAGCCAGCTGATCTACAAGTTCCTGAAG GCCGGCTACCTGGAGAACTGGCAGTACCACAAGACCTACAGCGGC3'; hLtrA-R3: 5'CACTTCATGTCCTTGATCTTCAGGTTGATCAGGCCGATCAGGGTCACGT-GGTCGATGTTGTCGAAGCAGCCCTTGATGTCGCCCTCCACGAACCACC3'; hLtrA-R4: 5'CAGCTGCAGCACGAACTTGTCCAACTCGTGCAGGTAGATGTTGG-CCAGCAGGGGGCTCAGGATGCCGCCCTGGGGGGTGCCGCTGTAGGTCTTGTG-GTACTGCC3'. Short oligonucleotides are s2-HIII: 5'AAAAAAGCTTGGAGTCCATC-TACGAGCCCGTGTTCGAGGACGTGAGCCACGGCTTCC3' and s2-XbaI: 5'AAAA-TCTAGACTCGCGGTCGAACTTCATCTTCAGCTGCAGCACGAACTTGTCCAAC-TCG3'.

For segment 3, long oligonucleotides are hLtrA-F5: 5'CCCGAGCGCATC-ACCCCCGAGTACCGCGAGCTGCACAACGAGATCAAGCGCATCAGCCACCGCC TGAAGAAGCTGGAGGGCG3'; hLtrA-F6: 5'CCCCTGCACCAGCCAGACCAACA-AGGTGCTGAAGTACGTGCGCTACGCCGACGACTTCATCATCAGCGTGAAGG-GCAGCAAGGAGG3'; hLtrA-R5: 5'GTCTGGCTGGTGCAGGGGAGGGTGGGCAG-CCTCTTGCGCTTCTCCTGGTACTCCAGCAGGACCTTGGCCTTCTCCTCGCCCT-CCAGCTTCTTCAGGC3'. Short oligonucleotides are s3-HIII: 5'AAAAAAGCTTGTTCGACCGCGAGAGCCCC-GAGCGCATCACCCCCG3' and s3-XbaI: 5'AAAATCTAGACCTTGATCCACTGGCAGTCCTCCTTGCTGCCCTTCA-CGCTGATG3'.

For segment 4, long oligonucleotides are hLtrA-F7: 5'GCTGAAGCTGTTCATC-CACAACAAGCTGAAGATGGAGCTGAGCGAGGAGAAGACCCTGATCACCCAC-

AGCCAGCCCGCCCGCTTCCTGGGC3'; hLtrA-R6: 5'CCACGCTGCCGTTCAGGG-TGCGCTTCTTCACCTTGCCGCTGCGCTTGATGGTGCCGCTGCGGCGCACGCG-GATGTCGTAGCCCAGGAAGCGGGCGGGCTGG3'; hLtrA-F8: 5'GCGCACCCT-GAACGGCAGCGTGGAGCTGCTGATCCCCCTGCAGGACAAGATCCGCCAGTT-CATCTTCGACAAGAAGATCGCCATCCAGAAGAAGGACAGC3'. Short oligonucleotides are s4-HIII: 5'AAAAAAGCTTGCCAGTGGATCAAGGAGCAGCT-GAAGCTGTTCATCC-ACAACAAGC3' and s4-XbaI: 5'AAAATCTAGAGCGGTGC-ACGGGGAACCAGCTGCTGTCCTTCTTCTGGATGGCG3'.

For segment 5, long oligonucleotides are hLtrA-F9: CCTGATCCGCAGC-ACCGACCTGGAGATCATCACCATCTACAACAGCGAGCTGAGGGGCATCTGC-AACTACTACGGCCTGGCCAGCAACTTC3'; hLtrA-F10: 5'CATCGCCTCCAAGC-ACAAGGGCACCCTGAGCAAGACCATCTCCATGTTCAAGGACGGCAGCGGCA-GCTGGGGCATCCCCTACGAGATCAAGCAGGGC3'; hLtrA-R7: 5'GCCCTTGTGC-TTGGAGGCGATGGTCTTCAGGCAGCTGTACTCCATCAGGTAGGCGAAGTAGT-TCAGCTGGTTGAAGTTGCTGGCCAGGCCGTAGTAG3'; hLtrA-R8: 5'CGTCGGT-GAACTGGTAGGGGGACTTGCACTCGCTGAAGTTGGCGAAGTAGCGGCGCTG-CTTGCCCTGCTTGATCTCGTAGGGG3'. Short oligonucleotides are s5-HIII: 5'AAA-AAAGCTTCGTGCAC-CGCAAGTACCTGATCCGCAGCACCGACCTGG3' and s5-XbaI: 5'AAAATCTAGA-CAGCACGGGGGCCTGGCTGATCTCGTCGGTGAACTG-GTAGGGGG3'.

For segment 6, long oligonucleotides are hLtrA-F11: 5'CTGTACGGCTAC-GCCCGCAACACCCTGGAGAACCGCCTGAAGGCCAAGTGCTGCGAGCTGTGC-GGCACCAGCGACGAGAACACC3'; hLtrA-F12: 5'GGGCAAGGAGAAGTGGGAG-ATGGCCATGATCGCCAAGCAGCGCAAGACCCTGGTGGTGTGCTTCCACTGCC-ACCGCCACG-TGATCCAC3'; hLtrA-R9: 5'GGCCATCTCCCACTTCTCCTTGCCC-

TTCAGGTTCT-TCACCTTGTTCACGTGGTGGATCTCGTAGCTGGTGTTCTCGTC-GCTGGTGCCG-CAC3'. Short oligonucleotides are s6-HIII: 5'AAAAAAGCTTGCCA-GGCCCCCGTGCTGTACGGCTACGCCCGCAACACCC3' and s6-XbaI: 5'AAAATC-TAGAGTCGACCTTGTGC-TTGTGGATCACGTGGCGGTGGC3'. Oligonucleotides used for overlapping segments 5 and 6 were hLtrA5/6F: 5'GCCAGGCCCCCGTGCTG3' and hLtrA5/6R: 5'CAGCACGGG-GGCCTGGC3'.

For each segment, PCR was done with the long oligonucleotides as overlapping templates. To reduce mutations, PCR reactions were carried out by using Vent DNA polymerase (New England Biolabs, Boston, MA), high annealing temperatures (58-60°C), and manual hot start (adding Vent DNA polymerase after sample temperature reached 94°C). PCR products were gel purified and digested with *Eco*RI and *Xba*I or *Hind*III and *Xba*I, then cloned into pKSBluescript to form pKS-hLtrA1-pKS-hLtrA6. After sequencing, a correct clone of pKS-hLtrA1 was digested with *Eco*RI and *Xba*I first, gel purified and then digested with *Hinf*I. A correct clone of pKS-hLtrA2 was digested with *Hind*III and *Xba*I first, gel purified, and then digested with *Hinf*I. Digested segments 1 and 2 were then ligated to the vector pKSBluescript digested with *Eco*RI and *Xba*I to construct pKS-hLtrA1/2. Similarly, segments 3 and 4 were ligated at *Tsp*RI site to pKSBluescript digested with *Hind*III and *Xba*I to construct pKS-hLtrA3/4. Segments 5 and 6 were gel purified and used as overlapping templates to obtain a fragment flanked by *Hind*III and *Xba*I sites and cloned to the vector accordingly to form pKS-hLtrA5/6. pKS-hLtrA1/2 was then digested with *Eco*RI and *Xba*I, gel purified, and again digested with *Bsi*EI. pKS-hLtrA3/4 was then digested with *Hind*III and *Xba*I, gel purified, and then digested with *Bsi*EI. pKS-hLtrA(1-4) was constructed by ligating hLtrA1/2 and hLtrA3/4 at the *Bsi*EI site and to pKSBluescript at *Eco*RI and *Xba*I sites. Finally, pKS-hLtrA was constructed by ligating hLtrA (1-4) and hLtrA5/6 at the *Apa*LI site and to

pKSBluescript at the *Eco*RI and *Not*I sites. phLtrA was constructed by cutting and pasting the *Eco*RI and *Sal*I fragment from pKS-hLtrA and annealed oligonucleotides SV40-F and SV40-R, flanked by *Sal*I and *Not*I sites, to corresponding sites in pIRES. pLtrA was constructed by ligating the PCR-amplified bacterial version of LtrA ORF into the*Eco*RI and *Sal*I sites of phLtrA. ΔNLS was constructed by replacing the 1.8-kb *Eco*RI/*Sal*I fragment of phLtrA with a PCR-amplified 1.8-kb fragment from pELG2-ΔORF, flanked by the same restriction sites.

6.2.4 Western Analysis

Cells were collected 48 h after transfection. To achieve even loading, the same number of cells was used in each lane. The transferred membrane was stained with AuroDye after blotting to confirm even loading. Anti-LtrA antibody was used at 1:1000 dilution, and the secondary antibody was used at 1:60,000 dilution.

6.2.5 RT-PCR

Total RNAs were prepared 48-h post transfection with a QIA RNeasy mini kit (Qiagen, Inc.). RT-PCR was done using the first-strand reverse transcription kit from Invitrogen following the manufacturer's instructions. Primer 3'BamHI was used for RT, which has the following sequence: 5'-GTGGATCCCCCAGCTTTGCCGC-3'. Primer E1-F, with the following sequence: 5'-GAGAAAAATAATGCGGTGCTTGG-3', was used in combination with primer 3'BamHI in PCR.

6.2.6 Immunofluorescence

The entire procedure was carried out at room temperature. Cells were first washed with PBS twice, and then fixed in 2% paraformaldehyde for 30 min, at room temperature. After three washes with PBS, 0.5% Triton X-100 in PBS was used to permeabilize the cells for 15 min, followed by three washes with PBS containing 0.2% Tween 20 (PBST).

Blocking was achieved by incubating the permeabilized cells with 10% normal goat serum and 1% BSA in PBST for 1 h. Primary antibody was first incubated with untransfected cell lysate (prepared by sonication) to deplete nonspecific antibody and then was used at 1:500 dilution in blocking buffer for a 1 h incubation. After four five-min washes in PBST containing 0.1 M NaCl, cells were incubated with 1:100 dilution of goat anti-rabbit antibody conjugated with fluorescein in blocking buffer for 1 h, washed with PBST containing 0.1 M NaCl five times for five min each time, incubated with 2 μg/ml Hoechst for 10 min, and washed twice with PBS. After mounting, cells were observed under a fluorescence microscope.

6.2.7 Construction of stable cell lines expressing hLtrA

To construct stable cell lines expressing LtrA, phLtrA was first linearized with *Bam*HI, phenol extracted, and ethanol precipitated. The linearized phLtrA was then electroporated into the cells. Forty-eight hours after electroporation, neomycin was used to select for cells containing one or more integrated neomycin-resistance genes. Neomycin-sensitive cells started to die after two weeks of selection. Finally, neomycin-resistant colonies were picked up, seeded into 96-well dishes and grown under selection conditions. Western analysis was used to detect LtrA expression.

6.2.8 Purification of Ll.LtrB RNPs from *E. coli* and RNP reconstitution *in vitro*

RNPs were purified as described for LtrA protein synthesized from pImp-1p in Saldanha et al. (1999) with following modifications. The intron and LtrA sequences were cloned into expression vector, pImp-1p, so that both the intron RNA and the LtrA are under the control of T7 promoter. The C-terminus of LtrA is fused to the N-terminus of a chitin binding domain with an intein cleavage site and expressed downstream of the intron. The RNP contruct was transformed to E. coli BL21 (DE3), and a single colony

was used to start an overnight culture. The next day, cells were diluted 1:50 into fresh LB medium plus ampicillin and grown until $O.D._{595}$ equals 0.6, then induced with 0.5 mM IPTG for 3 h at 25°C. Cells were collected and lysed in column buffer (500 mM NaCl, 25 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) with 10 ug/ml of lysozyme and three cycles of freezing and thawing between –80°C and 25°C. The lysate was then loaded onto a chitin column for affinity purification. In order to minimize nuclease exposure, the flow rate was set at ≥ 1 mL/min. The loaded column was first washed with 100 mL of column buffer containing 750 mM NaCl, then with 500 mM NaCl. After overnight cleavage with 30 mM DTT, RNPs were eluted with two bed-volume of column buffer (500 mM NaCl) and ultracentrifuged overnight in the Beckman 50.2 Ti rotor at 50,000 rpm and 4°C. The pellet was washed in ice-cold water and resuspended in ice-cold 10 mM Tris-HCl, pH 8.0, 1 mM DTT. In vitro reconstitution of RNPs was done essentially as described in Saldanha et al. (1999) with an additional step of ultracentrifugation step as described above.

6.2.8 Electroporation of RNPs

Cells were grown to about 80% confluence, trypsinized, diluted with growth medium, and washed with ice-cold PBS at least 4 times. They were then resuspend in an appropriate volume of PBS to achieve a density of $5x10^6$/ml. Ten micrograms of RNPs were mixed with 400 µl of cells and incubated on ice for less than 2 min. Electroporations were carried out at 230 v and various capacitance settings. After electroporation, to minimize the degradation of the RNPs by nucleases in the medium, cuvettes were left on ice for 10-15 min before 1 ml of growth medium was added to the cuvettes and cells were aliquoted into two 60-mm dishes. Fluorescence microscopy, flow cytometry, and immunofluorescence were performed 24 h after electroporation.

6.2.9 Cell preparation for FACS analysis

Transfected cells were trypsinized and neutralized with growth medium. After washing in PBS, cells were resuspended in PBS and run through a FACS.

6.2.10 *In vitro* TPRT-PCR

Ten micrograms of chromosomal DNA and 1-10 µg RNPs were added to 20 µl reaction medium containing 40 mM Tris-HCl, pH 7.5, 10 mM KCl, 10 mM $MgCl_2$, and 0.2 mM dNTPs. The reaction was incubated at 37°C for 2 h, followed by PCR using one tenth of the TPRT reaction as template. For targeting in nuclear lysates, one to two microliter RNPs (2-5 µg/µl) and 2 µl 2 mM dNTPs were added to 15 or 17 µl lysates with or without 2 µl 10xRT buffer. After TPRT, each reaction was treated with proteinase K for > 1 h at 55°C and then heated to 98°C for 30 min to inactivate proteinase K prior to PCR.

For detecting the 3' integration junction, intron primers ΔDIV-F (5'CAGAGCCGTATACTCCGAGAGG3') and LtrBs1 (CAGTAATGTGAACAAGGCGGTACCT3') were used. For detecting the 5' junction, intron primers slx-seq (5'CTTGTTTAGGGTATCCCCAG3') and 400-R (AACCGAAATTAGAAACTTGCGTTC3') were used. The PCR program was 94°C, 3 min, 30 cycles of 94°C, 30s, 58°C, 30s, and 72°C, 1 min 30 s, and an additional extension at 72°C for 10 min. I found it helpful to include 2-4% DMSO in PCR reactions involving chromosomal DNA. DMSO lowers the melting temperature of nucleic acids with high GC content (Escara & Hutton, 1980).

6.2.11 Preparation of nuclear extract

HEK 293 cells were grown to full confluence, washed with PBS, and blown off the dishes with ice-cold hypotonic buffer (10 mM HEPES, 10 mM KCl, 1 ml/100 mm

dish), and incubated on ice for 15 min. Cells were broken by 15 strokes of Dounce homogenizer. Nuclei were collected by centrifugation at 800xg for 5 min at 4°C. The supernatant was then drained. Complete separation of the nucleus from the cytoplasm was not critical for my purposes. Nuclei were resuspended in the residual buffer in the tube. Typically, 4 to 5 100-mm dishes of cells made 100 to 150 µl of lysate. After 3 cycles of –80°C/25°C freezing and thawing, chromosomal DNA was sheared by repeated pipetting. Eight µl of the solution was used in each *in vitro* TPRT reaction.

6.2.12 Purification of chromosomal DNA from cultured cells

Most of the time, cells were pelleted and washed with PBS before subjected to proteinase K digestion at 55˚C water bath for overnight in a solution containing 50 mM Tris, pH 7.5, 0.5% SDS, 0.2 mg/ml proteinase K. The homogeneous solution was then phenol extracted, and the top phase was transferred to a new tube with 1/10 vol 5 M NH_4Cl and 1 vol isopropanol. Chromosomal DNA was then picked out with a tip after a few inversion of the tube, washed, dried, and dissolved in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). However, the quality of DNA was better when cells were first lyzed with cold lysis buffer (0.32 M sucrose, 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1% Triton X-100), and nuclei were collected before proteinase K digestion.

6.2.13 Staining of *lacZ* expressing cells

Cells were fixed as in section 6.2.6 and then washed in washing buffer (2 mM $MgCl_2$, 0.02% NP40, 0.1 M sodium phosphate buffer, pH 7.3) for three times at room temperature for 15 m each before subjected to overnight staining in washing buffer containing 1 mg/ml X-gal, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide). Cells were then kept in washing buffer at 4˚C.

6.2.14 Mg^{2+} treatment of the cells

$MgCl_2$ was either added onto the cells in medium to a final concentration of 80 mM or cells were fed with fresh medium containing 80 mM $MgCl_2$. The latter may avoid local unevenness. It may be more important for HEK 293 cells because they do not adhere well. Cells were left in 80 mM $MgCl_2$ for 1-4 h before harvesting for RNA preparation. In order to avoid self-splicing during RNA preparation, I washed the cells with 0.5 M EDTA prior to lysis.

References

Aizawa, Y., Xiang, Q., Lambowitz, A. M., & Pyle, A. M. (2003) The pathway for DNA recognition and RNA integration by a group II intron retrotransposon. *Mol. Cell*, **11**, 795-805.

Anziano, P. Q. & Butow, R. A. (1991). Splicing-defective mutants of the yeast mitochondrial *COXI* gene can be corrected by transformation with a hybrid maturase gene. *Proc. Natl. Acad. Sci. USA,* **88**, 5592-5596.

Ausubel, F. M., Brent, F., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (1994) Useful Measurements and Data in Current Protocols in Molecular Biology, pp. A.1.5. John Wiley & Sons, Inc.

Belfort, M., Derbyshire, V., Parker, M. M., Cousineau, B. & Lambowitz, A. M. (2002). Mobile introns: pathways and proteins. In *Mobile DNA II* (Craig, N. L., Craigie, R., Gellet, M. & Lambowitz, A. M., eds.), pp. 761-783. ASM press, Washington, DC.

Belfort, M., Ehrenman, K. & Chandry, P. S. (1990). Genetic and molecular analysis of RNA splicing in *Escherichia coli*. *Methods Enzymol.* **181**, 521-539.

Bertrand, E., Castanotto, D., Zhou, C., Carbonnelle, C., Lee, N.S., Good, P., Chatterjee, S., Grange, T., Pictat, R., Kohn, D., Engelke, D., & Rossi, J.J. (1997) The expression cassette determines the functional activity of ribozymes in mammalian cells by controlling their intracellular localization. *RNA*, **3**, 75-88.

Bibillo, A. & Eickbush, T. H. (2002). The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. *J. Mol. Biol.* **316**, 459-473.

Blocker, F. J., Mohr, G., Conlan, L. H., Qi, L., Belfort, M., & Lambowitz, A. M. (2005) Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA,* **11**, 14-28.

Bock-Taferner, P. & Wank, H. GAPDH enhances group II intron splicing in vitro. *Biol. Chem.* **385**, 615-21.

Bojanowski, K. & Ingber, D.E. (1998) Ionic control of chromosome architecture in living and permeabilized cells. *Exp. Cell Res.* **244**, 286-294.

Boudvillain, M. & Pyle, A. M. (1998) Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.* **17**, 7091-7104.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254.

Brouha, B, Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., Kazazian, H. H., Jr. (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. USA***, 100**, 5280-5285.

Cannon, J. P., Colicos, S. M., & Belmont, J. W. (1999) Gene trap screening using negative selection: identification of two tandem, differentially expressed loci with potential hematopoietic function. *Dev. Genet.* **25**, 49-63.

Carignani, G., Groudinsky, O., Frezza, D., Schiavon, E., Bergantino, E. & Slonimski, P. P. (1983). An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in *S. cerevisiae*. *Cell,* **35**, 733-742.

Chen, B. & Lambowitz, A. M. (1997). *De novo* and DNA primer-mediated initiation of cDNA synthesis by the Mauriceville retroplasmid reverse transcriptase involve recognition of a 3' CCA sequence. *J. Mol. Biol.* **271**, 311-332.

Chen, Z. J. and Pikaard, C. S. (1997) Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Gene Dev.* **11**, 2124-2136.

Chin, K. & Pyle, A. M. (1995) Branch-point attack in group II introns is a highly reversible transesterification, providing a potential proofreading mechanism for 5'-splice site selection. *RNA,* **1**, 391-406.

Chu, V. T., Adamidi, C., Liu, Q, Perlman, P. S., & Pyle, A. M. (2001) Control of branch-site choice by a group II intron. *EMBO J.* **20**, 6866-6876.

Cormack, B. P., Valdivia, R. H. & Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene,* **173**, 33-38.

Cousineau, B., Smith, D., Lawrence-Cavanagh, S., Mueller, J. E., Yang, J., Mills, D., Manias, D., Dunny, G., Lambowitz, A. M. & Belfort, M. (1998). Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell,* **94**, 451-462.

Cui, X., Matsuura, M., Wang, Q., Ma, H., & Lambowitz, A. M. (2004) A Group II Intron-encoded Maturase Functions Preferentially *In Cis* and Requires Both the Reverse Transcriptase and X Domains to Promote RNA Splicing. *J. Mol. Biol.* **340**, 211-231.

Curcio, M. J. & Belfort, M. (1996). Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. *Cell,* **84**, 9-12.

Dai, L., Toor, N., Olson, R., Keeping, A., & Zimmerly, S. (2003) Database for mobile group II introns. *Nuclei Acids Res.* **31**, 424-426.

Davies, R. W., Waring, R. B., Ray, J. A., & Scazzocchio, C. (1982) Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature*, **300**, 719-724.

Deminoff, S. J., Tornow, J. & Santangelo, G. M. (1995). Unigenic evolution: a novel genetic method localizes a putative leucine zipper that mediates dimerization of the *Saccharomyces cerevisiae* regulator Gcr1p. *Genetics,* **141**, 1263-1274.

Dickson, L., Connell, S., Huang, H. R., Henke, R. M., Liu, L., & Perlman, P. S. (2004) Abortive transposition by a group II intron in yeast mitochondria. *Genetics,* **168**, 77-87.

D'Souza, L. M. & Zhong, J. (2002). Mutations in the *Lactococcus lactis* Ll.LtrB group II intron that retain mobility *in vivo*. *BMC Mol. Biol*. **3**, 17.

Eickbush, T. H. (1992) Transposing without ends: the non-LTR retrotransposable elements. *New Biol.* **4**, 430-440.

Eickbush, T. H. (1994) Origin and evolutionary relationships of reteoelements. In *The Evolutionary Biology of Viruses*, Morse, S. S., Ed. (New York: Raven Press, Ltd.), pp.121-157.

Ems, S. C., Morden, C. W., Dixon, C. K., Wolfe, K. H., dePamphilis, C. W., & Palmer, J. D. (1995) Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*. *Plant Mol. Biol.* **29**, 721-733.

Eperon, I. C. & Krainer, A. R. (1993) Splicing of mRNA precursors in mammalian cells in RNA Processing, a practical approach, Volume I. Edited by Higgins, S. J. & Hames, B. D. IRL press at Oxford University Press, Oxford.

Epinat, J., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F., and Lacroix, E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucl. Acids Res.* **31**, 2952-2962.

Escara, J. F. & Hutton, J. R. (1980) Thermal stability and renaturation of DNA in dimethyl sulfoxide solutions: acceleration of the renaturation rate. *Biopolymers,* **19**, 1325-1327.

Eskes, R., Liu, L., Ma, H., Chao, M. Y., Dickson, L., Lambowitz, A. M., & Perlman, P. S. (2000) Multiple homing pathways used by yeast mitochondrial group II introns. *Mol. Cell Biol.* **20**, 8432-8446.

Eskes, R., Yang, J., Lambowitz, A. M. & Perlman, P. S. (1997). Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell,* **88**, 865-874.

Fedoroya, O., Mitros, T., & Pyle, A. M. (2003) Domains 2 and 3 interact to form critical elements of the group II intron active site. *J. Mol. Biol.* **330**, 197-209.

Feng, Q., Moran, J. V., Kazazian, H. H. Jr., & Boeke, J. D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell,* **87**, 905-916.

Fire, A., Xu, S., Montgomery, M.K., Kostas S.A., Driver S.E., & Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, **391**, 806-11.

Frazier, C. L., San Filippo, J., Lambowitz, A. M., and Mills, D. A. (2003) Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. *Appl. Environ. Microbiol.* **69**, 1121-1128.

Fromant, M., Blanquet, S. & Plateau, P. (1995). Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. *Anal. Biochem.* **224**, 347-353.

Fu, D. & Collins, K. (2003) Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs. *Mol. Cell*, **11**, 1361-1372.

Fujimoto, H., Hirukawa, Y., Tani, H., Matsuura, Y., Hashido, K., Tsuchida, K., Takada, N., Kobayashi, M., & Maekawa, H. (2004) Integration of the 5' end of the retrotransposon, R2Bm, can be complemented by homologous recombination. *Nucl. Acids Res.* **32**, 1555-1565.

Garber, P. M., Vidanes, G. M., & Toczyski, D. P. (2005) Damage in transition. *Trends Biochem. Sci*. **30**, 63-66.

Gencheva, M., Anachkova, B. and Russev, G. (1996) Mapping the sites of initiation of DNA replication in rat and human rRNA genes. *J. Biol. Chem.* **271**, 2608-2614.

Gordon, P. M., Sontheimer, E. J., & Piccirilli, J. A. (2000) Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA,* **6**, 199-205.

Green PJ, Pines O, Inouye M. (1986) The role of antisense RNA in gene regulation. *Annu. Rev. Biochem.* **55**, 569-597.

Gregan, J., Bui, D. M., Pillich, R., Fink, M., Zsurka, G., & Schweyen, R. J. (2001a) The mitochondrial inner membrane protein Lpe10p, a homologue of Mrs2p, is essential for magnesium homeostasis and group II intron splicing in yeast. *Mol. Gen. Genet.* **264**, 773-781.

Gregan, J., Kolisek, M., & Schweyen, R. J. (2001b) Mitochondrial Mg(2+) homeostasis is critical for group II intron splicing in vivo. *Genes Dev.* **15**, 2229-2237.

Guo, H., Karberg, M., Long, M., Jones, J. P. III., Sullenger, B. & Lambowitz, A. M. (2000). Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science,* **289**, 452-457.

Guo, H., Zimmerly, S., Perlman, P. S., & Lambowitz, A. M. (1997) Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J.* **16**, 6835-6848.

Guthrie, C. (1991) Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science,* **253**, 157-163.

Haas, J., Park, E., and Seed, B. (1996) Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Curr. Biol.* **6**, 315-324.

Hagmann, M., Adlkofer, K., Pfeiffer, P., Bruggmann, R., Georgiev, O., Rungger, D., & Schaffner, W. (1996) Dramatic changes in the ratio of homologous recombination to nonhomologous DNA-end joining in oocytes and early embryos of *Xenopus laevis. Biol. Chem. Hoppe Seyler.* **377**, 239-250.

Hetzer, M., Wurzer, G., Schweyen, R. J., & Mueller, M. W. (1997) Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature,* **386**, 417-420.

Horie, K., Nishiguchi, S., Maeda, D., Shimada, K. (1994) Structures of replacement vectors for efficient gene targeting. *J. Biochem.* (Tokyo). **115**, 477-485.

Huang, H. R., Chao, M. Y., Armstrong, B., Wang, Y., Lambowitz, A. M. & Perlman, P. S. (2003). The DIVa maturase binding site in the yeast group II intron aI2 is essential for intron homing but not for in vivo splicing. *Mol. Cell. Biol.* **23**, 8809-8819.

Huang, H. R., Rowe, C. E., Mohr, S., Jiang, Y., Lambowitz, A. M., & Perlman, P. S. (2005) The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc. Natl. Acad. Sci. USA*, **102**, 163-168.

Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science,* **282**, 1669-1675.

Ichiyanagi, K., Beauregard, A., Lawrence, S., Smith, D., Cousineau, B., & Belfort, M. (2002) Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol. Microbiol.* **46**, 1259-1272.

Isel, C., Westhof, E., Massire, C., Le Grice, S. F., Ehresmann, B., Ehresmann, C. & Marquet, R. (1999). Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J.* **18**, 1038-1048.

Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D. Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H. & Arnold, E. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA,* **90**, 6320-6324.

Jenkins, B. D., Kulhanek, D. J. & Barkan, A. (1997). Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell,* **9**, 283-296.

Karberg, M., Guo, H., Zhong, J., Coon, R., Perutka, J., and Lambowitz, A.M. (2001) Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat. Biotechnol.* **19**, 1162-1167.

Kazazian, H. H. Jr. (2004) Mobile elements: drivers of genome evolution. *Science,* **303**, 1626-1632.

Kennell, J. C., Moran, J. V., Perlman, P. S., Butow, R. A. & Lambowitz, A. M. (1993). Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell,* **73**, 133-146.

Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science,* **256**, 1783-1790.

Lambowitz, A. M. & Perlman, P. S. (1990). Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem. Sci.* **15**, 440-444.

Lambowitz, A. M. & Zimmerly, S. (2004) MOBILE GROUP II INTRONS. *Annu. Rev. Genet.* **38**, 1-35.

Lambowitz, A. M., Caprara, M. G., Zimmerly, S. & Perlman, P. S. (1999). Group I and Group II ribozymes as RNPs: clues to the past and guides to the future. In *The RNA World* (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds.), 2nd edit., pp. 451-485, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Le Hir, H., Nott, A., and Moore, M. J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.* **28**, 215-220.

Li, S., Xu, M., & Coelingh, K. (1995) Electroporation of influenza virus ribonucleoprotein complexes for rescue of the nucleoprotein and matrix genes. *Virus Res.* **37**, 153-161.

Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V., & Cech, T., R. (1997) Reverse transcriptase motifs in the catalytic subunit of telomerase.. *Science,* **276**, 561-567.

Luan, D. D., Korman, M. H., Jakubczak, J. K., & Eickbush, T. H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell,* **72**, 595-605.

Mansour, S. L., Thomas, K. R., & Capecchi, M. R. (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a geneal strategy for targeting mutations to non-selectable genes. Nature, 336: 348-352.

Matsuura, M., Noah, J. W. & Lambowitz, A. M. (2001). Mechanism of maturase-promoted group II intron splicing. *EMBO J.* **20**, 7259-7270.

Matsuura, M., Saldanha, R., Ma, H., Wank, H., Yang, J., Mohr, G., Cavanagh, S., Dunny, G. M., Belfort, M. & Lambowitz, A. M. (1997). A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.* **11**, 2910-2924.

Michel, F. & Ferat, J. (1995). Structure and activity of group II introns. *Annu. Rev. Biochem.* **64**, 435-461.

Michel, F., Jacquier, A., & Dujon, B. (1982) Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie.* **64**, 867-881.

Michel, F., Umesono, K., & Ozeki, H. (1989) Comparative and functional anatomy of group II catalytic introns--a review. *Gene,* **82**, 5-30.

Mills, D. A., Choi, C. K., Dunny, G. M. & McKay, L. L. (1994). Genetic analysis of regions of the *Lactococcus lactis* subsp. *lactis* plasmid pRS01 involved in conjugative transfer. *App. Environ. Microbiol.* **60**, 4413-4420.

Mohr, G. & Lambowitz, A. M. (2003). Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucl. Acids Res.* **31**, 647-652.

Mohr, G., Caprara, M. G., Guo, Q. & Lambowitz, A. M. (1994). A tyrosyl-tRNA synthetase can function similarly to an RNA structure in the *Tetrahymena* ribozyme. *Nature,* **370**, 147-150, 1994.

Mohr, G., Perlman, P. S. & Lambowitz, A. M. (1993). Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucl. Acids Res.* **21**, 4991-4997.

Mohr, G., Smith, D., Belfort, M., & Lambowitz, A. M. (2000) Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev.* **14**, 559-73.

Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J. Boeke, J. D., & Kazazian, H. H. Jr. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell,* **87**, 917-927.

Moran, J. V., Mecklenburg, K. L., Sass, P., Belcher, S. M., Mahnke, D., Lewin, A. & Perlman, P. (1994). Splicing defective mutants of the *COXI* gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron AI2. *Nucl. Acids Res.* **22**, 2057-2064.

Moran, J. V., Zimmerly, S., Eskes, R., Kennell, J. C., Lambowitz, A. M., Butow, R. A. & Perlman, P. S. (1995). Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol. Cell. Biol.* **15**, 2828-2838.

Morrish, T. A., Gilbert, N., Myers, J.S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., Batzer, M. A., & Moran, J. V. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* **31**, 159-165.

Morris, M. C., Depollier, J., Mery, J., Heitz, F., Heitz, F., & Divita, G. (2001) A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nat. Biotechnol.* **19**, 1173-1176.

Munoz-Adelantado, E., San Filippo, J., Martinez-Abarca, F., Garcia-Rodriguez, F. M., Lambowitz, A. M., & Toro, N. (2003) Mobility of the *Sinorhizobium meliloti* group II intron RmInt1 occurs by reverse splicing into DNA, but requires an unknown reverse transcriptase priming mechanism. *J. Mol. Biol.* **327**, 931-943.

Neumann G, Watanabe T, Ito H, Watanabe S, Goto H, Gao P, Hughes M, Perez DR, Donis R, Hoffmann E, Hobom G, Kawaoka Y. (1999) Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl. Acad. Sci. USA,* **96**, 9345-9350.

Noah, J. W. & Lambowitz, A. M. (2003). Effects of maturase binding and Mg^{2+} concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry,* **42**, 12466-12480.

Olson, M. J. (1994) Introduction to nucleolus in The Nucleolus. Landes Biosciences/Eureckah.com, Georgetown, TX.

Ostheimer, G. J., Williams-Carrier, R., Belcher, S., Osborne, E., Gierke, J. & Barkan, A. (2003). Group II intron splicing factors derived by diversification of an ancient RNA-binding domain. *EMBO J.* **22**, 3919-3929.

Palmer, T. D., Miller, A. D., Reeder, R. H., and McStay, B. (1993) Efficient expression of a protein coding gene under the control of an RNA polymerase I promoter. *Nucl. Acids. Res.* **21**, 3451-3457.

Perron, K., Goldschmidt-Clermont, M. & Rochaix, J. (1999). A factor related to pseudouridine synthases is required for chloroplast group II intron *trans*-splicing in *Chlamydomonas reinhardtii*. *EMBO J.* **18**, 6481-6490.

Perron, K., Goldschmidt-Clermont, M., & Rochaix, J. D. (2004) A multiprotein complex involved in chloroplast group II intron splicing. *RNA,* **10**, 704-711.

Perutka, J., Wang, W., Goerlitz, D., and Lambowitz, A.M. (2004) Use of computer-designed group II introns to disrupt Escherichia coli DexH/D-box protein and DNA helicase genes. *J. Mol. Biol.* **336**, 421-439.

Pietrobono, R., Pomponi, M. G., Tabolacci, E., Oostra, B., Chiurazzi, P, and Neri, G. (2002) Quantitative analysis of DNA demethylation and transcriptional reactivation of the FMR1 gene in fragile X cells treated with 5-azadeoxycytidine. *Nucl. Acids Res*. **30**, 3278-3285.

Podar, M., Chu, V. T., Pyle, A. M., & Perlman, P. S. (1998) Group II intron splicing in vivo by first-step hydrolysis. *Nature,* **391**, 915-918.

Pyle, A.M., & Lambowitz, A.M. Group II introns: ribozymes that splice RNA and invade DNA. In: *The RNA World*, Third Edition (R.F. Gesteland, T.R. Cech, and J.F. Atkins, Editors), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, in press.

Qin, P. Z. & Pyle, A. M. (1998) The architectural organization and mechanistic function of group II intron structural elements. *Curr. Opin. Struct. Biol.* **8**, 301-308.

Reznikoff, W. S., Goryshin, I. Y., & Jendrisak, J. J. (2004) Tn5 as a Molecular Genetics Tool: In Vitro Transposition and the Coupling of In Vitro Technologies with in Vivo Transposition. *Methods Mol. Biol.* **260**, 83-96.

Romani, A. M. & Scarpa, A. (2000) Regulation of cellular magnesium. *Front Biosci.* **5**, D720-734.

Rivier, C., Goldschmidt-Clermont, M., & Rochaix, J. D. (2001) Identification of an RNA-protein complex involved in chloroplast group II intron trans-splicing in *Chlamydomonas reinhardtii*. *EMBO J.* **20**, 1765-1773.

Robart, A. R., Montgomery, N. K., Smith, K. L., & Zimmerly, S. (2004) Principles of 3' splice site selection and alternative splicing for an unusual group II intron from Bacillus anthracis. *RNA,* **10**, 854-862.

Rubnitz, J. & Subramani, S. (1984) The Minimum Amount of Homology Required for Homologous Recombination in Mammalian Cells. *Mol. Cell Biol.* **4**, 2253-2258.

Ruskin, B. & Green, M. R. (1985) An RNA processing activity that debranches RNA lariats. *Science*, **229**, 135-140.

Saldanha, R., Chen, B., Wank, H., Matsuura, M., Edwards, J. & Lambowitz, A. M. (1999). RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry,* **38**, 9069-9083.

San Filippo, J. & Lambowitz, A. M. (2002). Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *J. Mol. Biol.* **324**, 933-951.

Sashital, D. G., Cornilescu, G., & Butcher, S. E. (2004) U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat. Struct. Mol. Biol.* **11**, 1237-1242.

Schemlzer, C., Schmidt, C., & Schweyen, R. J. (1982) Identification of splicing signals in introns of yeast mitochondrial split genes: mutational alterations in intron bI1 and secondary structures in related introns. *Nucl. Acids Res.* **10**, 6797-6808

Séraphin, B., Simon, M., Boulet, A. & Faye, G. (1989). Mitochondrial splicing requires a protein from a novel helicase family. *Nature,* **337**, 84-87.

Shukla, F. C. & Padgett, R. A. (2002) A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome.*Mol. Cell,* **9**, 1145-1150.

Singh, N. N. & Lambowitz, A. M. (2001) Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J. Mol. Biol.* **309**, 361-386.

Singh, R. N., Saldanha, R. J., D'Souza, L. M. & Lambowitz, A. M. (2002). Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J. Mol. Biol.* **318**, 287-303.

Smale, S. T. & Tjian, R. (1985) Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters. *Mol. Cell Biol.* **5**, 352-62.

Sontheimer, E. J., Gordon, P. M., &Piccirilli, J. A. (1999) Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev.* **13**, 1729-1741.

Spaete, R. R. & Mocarski, E. S. (1985) Cytomegalovirus Gene Expression: α and β Promoters Are *trans* Activated by Viral Functions in Permissive Human Fibroblasts. *J. Virol.* **56**, 135-143.

Steitz, T. A. & Steitz, J. A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA***, 90**, 6498-6502.

Thomas KR, Capecchi MR. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, **51**,503-512.

Till, B., Schmitz-Linneweber, C., Williams-Carrier, R. & Barkan, A. (2001). CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. *RNA,* **7**, 1227-1238.

Toor, N., Hausner, G. & Zimmerly, S. (2001). Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA,* **7**, 1142-1152.

Valadkhan, S. & Manley, J. L. (2003) Characterization of the catalytic activity of U2 and U6 snRNAs. *RNA,* **9**, 892-904.

Villa, T., Pleiss, J. A., & Guthrie, C. (2002) Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell,* **109**, 149-152.

Vogel, J. & Borner, T. (2002) Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *EMBO J.* **21**, 3794-3803.

Vogel, J., Börner, T. & Hess, W. R. (1999). Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucl. Acids Res.* **27**, 3866-3874.

Vogel, J., Hübschmann, T., Börner, T. & Hess, W. R. (1997). Splicing and intron-internal RNA editing of *trnK-matK* transcripts in barley plastids: support for MatK as an essential splice factor. *J. Mol. Biol.* **270**, 179-187.

Wank, H., SanFilippo, J., Singh, R. N., Matsuura, M. & Lambowitz, A. M. (1999). A reverse transcriptase/maturases promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol. Cell,* **4**, 239-250.

Watanabe, K. & Lambowitz, A. M. (2004) High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA. *RNA*, **10**, 1433-1443.

Webb, A. E., Rose, M. A., Westhof, E. & Weeks, K. M. (2001). Protein-dependent transition states for ribonucleoprotein assembly. *J. Mol. Biol.* **309**, 1087-1100.

Will, C. L. & Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290-301.

Xiong, Y. & Eichbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**, 3353-3362.

Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. (1996). Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature,* **381**, 332-335.

Yoon, Y., Sanchez, J. A., Brun, C., and Huberman, J. A. (1995) Mapping of replication initiation sites in human ribosomal DNA by nascent-strand abundance analysis. *Mol. Cell. Biol.* **15**, 2482-2489.

Young, N. D. & dePamphilis, C. W. (2000). Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. *Mol. Biol. Evol.* **17**, 1933-1941.

Yu, Y. T., Maroney, P. A., Darzynkiwicz, E., & Nilsen, T. W. (1995) U6 snRNA function in nuclear pre-mRNA splicing: a phosphorothioate interference analysis of the U6 phosphate backbone. *RNA,* **1**, 46-54.

Zhong, J. & Lambowitz, A. M. (2003) Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J.* **22**, 4555-4565.

Zhong, J., Karberg, M., and Lambowitz, A.M. (2003) Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucl. Acids Res.* **31**, 1656-1664.

Zhou, L., Manias, D. A. & Dunny, G. M. (2000). Regulation of intron function: efficient splicing *in vivo* of a bacterial group II intron requires a functional promoter within the intron. *Mol. Microbiol.* **37**, 639-651.

Zhu, W. & Brendel, V. (2003) Identification, characterization and molecular phylogeny of U12-dependent introns in the *Arabidoposis thaliana* genome. *Nucl. Acids Res.* **31**, 4561-4572.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P. S. & Lambowitz, A. M. (1995). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell,* **83**, 529-538.

Zimmerly, S., Guo, H., Perlman, P. S. & Lambowitz, A. M. (1995). Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell,* **82**, 545-554.

Zimmerly, S., Hausner, G. & Wu, X. (2001). Phylogenetic relationships among group II intron ORFs. *Nucl. Acids Res.* **29**, 1238-1250.

Vita

Xiaoxia Cui was born in Guantao, Hebei, China on November 25, 1971, to father, Zhizhong Cui, and mother, Xiue Miao. She attended Yangzhou Middle School (Yangzhou, Jiangsu, China) before enrolling in Nanjing University (Nanjing, China), where she received a bachelor's degree in Chemistry and then spent two years there in graduate studies in Biochemistry. She won the best class president award in 1992. In 1995, she joined the graduate program in the Department of Biochemistry at the University of Alabama at Birmingham and received a master's degree in basic medical sciences in 1997, before she moved to Austin. She has worked in Dr. Lambowitz's laboratory since then. She won the A.P. Bradie Endowed Fellowship Award for research excellence in 2004. She is a coinventor of two provisional patent applications. She coauthored the following papers:

Cui, X., Matsuura, M., Wang, Q, Ma, H., & Lambowitz, A.M. (2004) A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing. *J. Mol. Biol.* **340**, 211-231.

Alb, J.G.Jr., Phillips, S.E., Rostand, K., **Cui, X.**, Pinxteren, J., Cotlin, L., Manning, T., Guo, S., York, J.D., Sontheimer, H., Collawn, J.F., & Bankaitis, V.A. (2002) Genetic ablation of phosphatidylinositol transfer protein functionin murine embryonic stem cells. *Mol. Biol. Cell*, **13**, 739-754.

Li, X., Routt, S.M., Xie, Z., **Cui, X.**, Fang, M., Kearns, M.A., Bard, M., Kirsch, D.R., Bankaitis, V.A. (2000) Identification of a novel family of nonclassic yeast phosphatidylinositol transfer proteins whose function modulates phospho;ipase D activity and Sec14p-independent cell growth. *Mol. Biol. Cell*, **11**, 1989-2005.

Chen, H., Yan, H., Luo, L., **Cui, X**., and Tang, W. (1997) Study of the thermo decomposition, axial-base equilibrium, and Co-C bond dissociation energy of 2',

5’-dideoxyadenosylcobalamin in ethylene glycol. *J. Inorgan. Biochem.* **66**, 219-225.

Permanent address: 15308 Cadoz Dr., Austin, TX 78728

This dissertation was typed by Xiaoxia Cui.